

Pre-Print

Describing language assessments for school-aged children: A Delphi study

Deborah Denman^{1*}, Jae-Hyun Kim^{2, 1}, Natalie Munro^{3,1}, Renée Speyer^{4, 1, 5}, Reinie Cordier¹

¹ Faculty of Health Sciences, Curtin University, Perth, Australia

² Department of Linguistics, Macquarie University, Sydney, Australia

³ Faculty of Health Sciences, The University of Sydney, Sydney, Australia

⁴ Department of Special Needs Education, University of Oslo, Oslo, Norway

⁵ Department of Otorhinolaryngology and Head and Neck Surgery, Leiden University

Medical Centre, Leiden, The Netherlands

* Corresponding author

E-mail: deborah.denman@postgrad.curtin.edu.au

Short title: Describing Language Assessments for Children

Keywords: developmental language disorder, language impairment, assessment, terminology,

Delphi study

Abstract

Purpose:

Given the barriers that inconsistent terminology poses for the Speech-Language Pathology (SLP) profession, this study aimed to develop an agreed upon taxonomy with well-defined categories for describing language assessment practices for children.

Method:

A taxonomy with illustrative terms for describing assessments across four aspects (modality/domain, purpose, delivery, and form) was developed with reference to contemporary literature. In a three round Delphi study, SLPs with expertise in child language were asked to indicate their level of agreement with the taxonomy and provide feedback. Participants were also asked to apply the taxonomy by categorising assessments presented in case studies.

Results:

A total of 55 participants completed round one, while 43 and 32 completed rounds two and three respectively. Agreed consensus with the taxonomy was achieved in both rounds one and two, with at least 88% of participants agreeing with each aspect and 100% agreeing with the overall structure. In round three, agreement was reached on 7/10 components for one case study and 4/10 for the other.

Conclusion:

The development of this taxonomy represents a significant step towards providing detailed terminology for describing language assessments. Future research is needed to investigate implementation strategies to facilitate consistent application of the taxonomy by SLPs.

Background

Internationally, the problem of inconsistent use of professional terminology by speech language pathologists (SLPs) is widely recognised (Walsh & IGOTF-CSD., 2006). Often one term may be used to refer to a range of different concepts or, conversely, different concepts are described by the same term (Walsh, 2005). Inconsistently applied terminology leads to breakdowns in professional communication and thus limits scientific debate needed to advance the profession. Lack of detailed terminology also hinders research translation as practices may not be described well enough to be replicated (Roulstone, 2015; Walsh & IGOTF-CSD., 2006).

In the field of child language assessment, many models and terms exist for describing the different types of language assessments that SLPs may use. As a result, the way in which one SLP conceptualises and describes their language assessment may well be different to another SLP's description of the same assessment. This creates significant barriers for the collection of accurate data on current practice both within and across service agencies (Cowie et al., 2001). Without an accurate understanding of current SLP assessment practice, it is difficult to compare current practice with evidence-based practice and thus identify clinical recommendations that align contemporary practice with policy and research evidence (Eadie, 2003).

A framework frequently used to describe SLP assessment practice is the International Classification of Functioning and Disability (ICF) (McLeod & Threats, 2008; World Health Organisation, 2015). This framework was designed to provide a structure by which concepts related to health and well-being may be viewed but, as such, lacks detail for describing language assessment (Barnes & Bloch, 2018; Hughes & Orange, 2007). Since

communication spans all aspects of health and well-being, it is acknowledged that SLPs often experience difficulty mapping assessment practices onto ICF categories (Barnes & Bloch, 2018; Hughes & Orange, 2007). Considerable disparity exists across literature with regards to how language assessments are classified within the ICF. For example, in some studies, norm-referenced language assessments, such as editions of the Clinical Evaluation of Language Fundamentals (Wiig, Semel, & Secord, 2004) are identified as assessing the body structure and functioning category of the ICF, while other studies identify these measures as assessing the activity category of the ICF (Paul & Norbury, 2012; Westby & Washington, 2017). The development of frameworks that are specifically targeted at describing SLP practices may facilitate greater consistency with regards to how assessments are described and thus enhance professional communication (Barnes & Bloch, 2018).

Specifically within the field of paediatric SLP, there are a number of terms used to describe language assessments. One common feature is to describe the skill domain targeted in the assessment. This may be through the use of Bloom and Lahey's taxonomy, which describes language domains across three aspects including form, content and use (Lahey, 1988); or through terms such as semantics, syntax, morphology, narrative or executive functioning (Larson & McKinley, 2007; Paul & Norbury, 2012). A second way in which assessments may be described is according to the purpose of the assessment. Categories include analytical or prognostic; summative or formative; or distinctions related to diagnostic purposes, screening, selecting intervention or determining service eligibility (Dockrell & Marshall, 2015; Newton, 2007; Paul & Norbury, 2012; Wade, 2004). Assessments may also be described by the method in which the assessment is conducted or the environmental context targeted in the assessment. Examples of methods include parent questionnaires, tests administered either face-to-face or via telehealth, or assessments conducted by automated

computer software. Examples of terms related to environmental context include curriculum-based, naturalistic or authentic (Parsons, Law, & Gascoigne, 2005; Paul & Norbury, 2012). Finally, assessments may be identified by the type of data collected or the type of tasks embedded in the assessment. This includes terms such as norm-referenced, criterion referenced and dynamic for describing type of data collected; or terms such as discrete-skill, functional, contextualised or language sampling for describing the types of tasks being assessed (American Speech and Hearing Association, 2018; Kaderavek, 2015).

In addition to the presence of numerous sets of terms for conceptualising language assessments, the definitions of these terms are often not precisely defined or are defined differently across literature. For example, the term standardised has been used to describe any assessment that has structured guidelines for administration (regardless of the type of data collected), but has also been used interchangeably with the term norm-referenced to refer specifically to assessments that provide normative data from a sample of age-matched peers (Kaderavek, 2015). Terminology used to describe assessments that are non-standardised in nature is even more loosely defined, with terms such as authentic, alternative, informal, naturalistic, behavioural and observational all used with unrestrained boundaries for the types of assessments covered by these terms (Caesar & Kohler, 2009; Hegde & Pomaville, 2017). Furthermore, detail in assessment practice is not captured through the use of one framework or one set of terms. Two assessments described by the same term could be vastly different. For instance, a morphology assessment could refer to a series of clinician directed sentence completion tasks organised developmentally or an analysis of the morphological forms produced in a language sample (Paul & Norbury, 2012). Similarly, a language sample might be a highly structured, norm-referenced narrative retell task or observations by an SLP during unstructured free play (Westerveld & Claessen, 2014). To describe assessments in detail,

SLPs need access to a framework that facilitates the conceptualisation of language assessments across multiple distinguishing features.

Given the current problems associated with terminology, there is a pressing need for actions that facilitate rigour and consistency with regards to the terms SLPs use for describing child language assessment (Walsh & IGOTF-CSD., 2006). It is evident that a single framework or a list of terms is unlikely to solve all problems related to such a complex problem (Walsh, 2005). Nonetheless, solutions are needed for situations where terminology must be consistently applied in order to be useful, such as when collecting survey data on the types of practices SLPs use (Cowie et al., 2001). With this in mind, the establishment of an agreed-upon taxonomy for conceptualising various child language assessment practices is a logical step towards addressing some of the challenges associated with inconsistent terminology. In addition to facilitating data collection, such a taxonomy has the potential to stimulate much-needed professional discussion and reflection on assessment practice, which is vital for continued advancement in the field (Eadie, 2003; Roulstone, 2001). There is also a significant need for further research examining the application of professional terminology. This will assist in better understanding the issues and complexities associated with developing consistent use of terminology in the SLP field (Walsh & IGOTF-CSD., 2006).

The current study

This study employed a Delphi study technique to address two aims: (1) to develop a taxonomy (i.e. categorisation framework) that is agreed upon by experts and provides distinct, well-defined categories for describing language assessment practises for children aged 4-18 years; and (2) to examine SLP application of a taxonomy for describing language assessments in clinical contexts. For the purposes of this study, *language assessment* may be

any data-gathering action including case histories, test performance, language sampling, behavioural observations, reports from significant others, and reports on educational achievement (Paul & Norbury, 2012). Given this is the first study to examine terminology for describing SLP assessment practice and acknowledging the potential complexities associated with developing consensus, the participant group in this study focussed on SLPs in a single country (Australia). It is envisaged that outcomes from this initial study will then act as the basis for further future research internationally.

Method

This study used a Delphi study technique with mixed-methods data collection and analysis (Tapio, Paloniemi, Varho, & Vinnari, 2011). The Delphi technique is a structured process which aims to develop group consensus on a defined topic through a series of survey rounds (Boulkedid, Abdoul, Loustau, Sibony, & Alberti, 2011). The same participants complete each round (though not all may continue with each round) and rounds are held until consensus is reached (or it becomes apparent that consensus cannot be reached). Results from previous rounds are used to inform changes that may facilitate consensus in subsequent rounds.

Participants

Criteria for participation in the Delphi study were: (1) eligibility for certified practicing membership with the Australian professional association for SLPs (Speech Pathology Australia); and (2) having spent at least 5 years (full-time equivalent) in the last 10 years engaged in professional activities, where 50% or more of professional time is related to children aged 4-18 years with a language support needs. These activities may include: research, academic teaching, consultancy, resource development, provision of SLP professional development, provision of clinical services or a combination of these activities.

Potential participants were contacted by email after being identified from the Speech Pathology Australia Find a Speech Pathologist website, the 2016 Speech Pathology Australia National Conference attendance contact list, and from the professional networks of the authors. In some states, recruited participants were also asked to identify other potential participants.

A total of 202 invitations were emailed and all SLPs who responded to invitations were sent a link to the first survey. As each survey was developed based upon the results of the preceding Delphi round, participants who did not complete a survey round were excluded from subsequent rounds. This helped to ensure that all participants had the same knowledge of the taxonomy. The number of participants who completed each Delphi round was 55 in round one (71.4% response rate), 43 in round two (78.2% response rate) and 32 in round three (74.4% response rate). Participant demographics for each round are presented in Table I.

<Insert Table I about here>

Procedure

Following a deductive (top-down) approach (DeJong, Horn, Gassaway, Slavin, & Dijkers, 2004), theoretical literature and research publications related to language assessment for school-aged children were reviewed by the first author to identify key concepts and terms that provide both a description of and differentiation between assessments. These concepts and terms were further developed through discussions within the research team and organised into an initial taxonomy. This taxonomy consisted of four aspects for describing features of assessments including: language modalities and domains, purposes, delivery methods and contexts, and the assessment form in terms of type of data collected and type of tasks used.

The initial taxonomy was then presented to study participants in a three round Delphi study for feedback. Each Delphi round was conducted as an online survey using Qualtrics software. The round one survey was piloted initially with two SLPs to check clarity of questions and completion time before being opened to Delphi study participants. Each survey was estimated as taking 90 minutes to complete. Delphi rounds were conducted between April-October 2017 with each survey being accessible for three to seven weeks. The study details were outlined at the beginning of each survey; participants were required to indicate consent to participate before accessing the remainder of the survey content. The study was ethically approved by the Curtin University Human Research Ethics Committee (Approval number: HRE2017-0126).

Each Delphi study round consisted of two parts. *Part A* addressed the first research aim of developing consensus regarding the structure and definitions of the taxonomy. These questions were Likert scale responses (Strongly Disagree, Disagree, Neither Agree or Disagree, Agree and Strongly Agree). Participants who did not indicate agreement with an aspect were asked an open-ended question about what they would change with regards to the structure or definitions within the aspect. *Part B* explored the second research aim of examining application of the taxonomy by SLPs when describing assessments. Participants were asked to select taxonomy categories that they thought applied to assessments presented in the Delphi study, with open-ended questions also provided for participants to make comments regarding the use of the taxonomy. In the first survey round, participants were asked to describe four assessments that were identified to them by name. These included: Clinical Evaluation of Language Fundamentals -4th Edition (CELF-4) (Wiig et al., 2004), Children's Communication Checklist – 2nd Edition (CCC-2) (Bishop, 2003), Language Sampling Protocol (Westerveld & Gillon, 2002), and the Pragmatics Profile of Everyday

Communication Skills in Children (Dewart & Summers, 1995). Participants who identified themselves as being unfamiliar with any of the assessments were not required to provide categorisations for those assessments. At least 24 of the study participants categorised each assessment for each aspect of the taxonomy.

Analysis of data after round one led to the use of case studies in survey rounds two and three. The use of case studies made it possible for all participants to complete all questions as background information was provided about each assessment as well as a link to the published test website. The case studies were constructed to examine components of the taxonomy that may be more difficult to apply, while still being assessments that were characteristic of paediatric SLP practice. Two assessments were embedded within the case studies and the same case studies were used in both rounds two and three. Case study one described a parent interview using the Pragmatic Profile of Everyday Communication Skills for Children (Dewart & Summers, 1995) for a 4;10 year old child with Autism Spectrum Disorder. Case study two described a language sampling procedure using the Language Sampling Protocol (Westerveld & Gillon, 2002) with a 7;10 year old child experiencing language difficulties at school. These assessments were selected as they had the greatest inconsistency in agreement noted in round one compared to the agreement for the CELF-4 (Wiig et al., 2004) and CCC-2 (Bishop, 2003)

In each of three survey rounds, the proposed taxonomy structure and definitions were presented in a reference document along with a summary of background information and references to literature. After each round, changes to the taxonomy in response to quantitative data (level of agreement with taxonomy structure and definitions or level of agreement with assessment categorisations) and qualitative data (themes from participant comments and

feedback) were made by updating the taxonomy reference document. Changes were made with the aim of either increasing agreement with the taxonomy itself, or improving application of the taxonomy by addressing identified sources of confusion with definitions. In rounds two and three, participants were also provided with a document summarising the quantitative and qualitative group results from the previous round.

Analysis

Survey responses were analysed using the Statistical Package for the Social Sciences (SPSS version 20 software (IBM Corp, Released 2011)). The number of rounds and criteria for agreed consensus were determined before the study commenced. In *Part A*, agreement with taxonomy structure and definitions was defined as 75% or more participants selecting “Strongly Agree” or “Agree” (i.e. median score of 4 or more on the five-point scale and inter-quartile range (IQR) of 1). In *Part B*, consensus on the categorisation of assessments was considered achieved when 75% or more participants selected (or opted not to select) a taxonomy category for an assessment. While agreement between the author’s categorisations and Delphi participant’s categorisations was not a requirement for consensus, examining concordance between the two provided an additional means of examining application of the taxonomy. Participant’s survey responses to open-ended questions were analysed using conventional content analysis (Hsiu-Fang & Shannon, 2005) to identify themes for each aspect of the taxonomy. These themes were considered alongside quantitative data to identify potential reasons for lack of participant consensus (Tapio et al., 2011). Data analysis was conducted by the first author, who was blinded to the identity of participants during analysis, and results were reviewed by the other authors.

Results

Part A: Agreement with taxonomy structure and definitions

Delphi participants reached consensus on the structure and definitions of the taxonomy in both rounds one and two, with 100% of participants strongly agreeing or agreeing with the overall structure of the taxonomy and at least 88.4% of participants strongly agreeing or agreeing with the sub-components and definitions within every aspect. No participants strongly disagreed with any aspect of the taxonomy. These results are provided in Table II. As consensus was established across both rounds one and two, participants were not asked to rate their level of agreement regarding the structure and definitions in round three. Therefore, the round three survey only contained content related to *Part B*.

<Insert Table II about here>

Part B: Categorisation of assessments using the taxonomy

At the end of round three, consensus was established regarding seven out of the 10 components for case study one (parent interview for a child with Autism Spectrum Disorder) and four out of the 10 components for case study two (language sampling for a child experiencing language difficulties at school). The level of agreement with regards to the categorisation of case studies across each taxonomy component is provided in Table III.

<Insert Table III about here>

Final Taxonomy

The agreed-upon taxonomy has four aspects, labelled in roman numerals I-IV, each containing a number of components that describe assessments. The finalised structure of the

taxonomy after round three is represented in Figure 1 and the finalised definitions of each taxonomy component after round three are provided in Supplementary Table I. Each taxonomy aspect is described below followed by a summary of the components that were not consistently categorised in case studies at the end of round three. The themes identified from participant comments and associated changes to the taxonomy through the Delphi study rounds are summarised in Supplementary Material II.

<Insert Figure 1 about here>

Aspect I (Modalities and Domains)

Aspect I provides terminology for describing the skills that are specifically measured by an assessment and reported on in assessment findings. There are three components: modality, domain and comprehension/production. Modality includes categories *spoken* and *written* (including AAC). Domains include *semantics*, *morphosyntax*, *social abilities & discourse*, *meta-abilities* and *executive functions*. Assessments are also described as targeting *comprehension* (reception) or *production* (expression) of language. The categories in Aspect I are not mutually exclusive. An assessment may target either or both *spoken* and *written* modalities, either or both *comprehension* and *production*, and target one or more domains.

At the end of round three, the categories *semantics* and *executive functioning* remained inconsistently selected. Themes from participant comments suggested the following possible reasons for lack of consensus: (1) participants considering other ways an assessment could be conducted or selecting domains that may be involved in completing assessment tasks, but are not the key domains being measured by the assessment; (2) perceived overlap between the

categories of *semantics* and *executive functioning*; and (3) the high level of information processing required from Delphi participants when reading and applying definitions.

Aspect II (Assessment Purpose)

Aspect II describes the purposes for which assessments are used. There are seven categories that include *predict outcome*, *select intervention*, *plan dosage* (prognostic purposes relating to identification of possible future needs or supports) and *screening*, *diagnostic*, *detect change* and *describe status* (analytical purposes related to describing current functioning). These categories are not mutually exclusive as an assessment may have more than one purpose.

After round three, consensus on all Aspect II categories was not reached for either case study. Themes from participants comments identified the following reasons for lack of consensus: (1) the possibility of participants selecting all possible ways an assessment could be used; (2) the potential for the *predict outcome* category to be only viewed as prognosis for improvement in diagnostic symptoms, rather than covering identification of future supports or need for intervention; and (3) individual SLP perceptions or service agency policy influencing categorisation. For example, if severity of diagnostic symptoms is used to determine eligibility for services within a particular clinical setting, then *diagnostic* purposes may not be differentiated from purposes of *predict outcome* or *select intervention*.

Aspect III (Service Delivery)

Aspect III provides terms for describing the methods and contexts in which assessments are conducted. This aspect has two components. The first component describes the method by which data is collected and includes three categories: (1) direct sampling, testing or observing a child's skills either by a SLP or by another trained person (e.g. teacher, parent or other

professional), (2) assessment administered through a software program; and (3) collection of proxy-reported information (e.g. getting information from a parent through an interview or checklist. Assessments conducted by a SLP or a trained person can occur either face-face or via telehealth using information and communication technologies (ICTs). The other component in Aspect III considers the *environmental context* targeted in the assessment. Assessment may occur within a clinical context or within *home, school* or *other* community contexts (Parsons et al., 2005; Schraeder, Quinn, Stockman, & Miller, 1999). Each category in Aspect III is mutually exclusive from other categories within each of the respective components (i.e. a particular assessment is conducted via only one method and targets only one *environmental context*).

After round three, lack of consensus remained with regards to the *environmental context* categories for both case studies. Participant comments identified: (1) lack of clarity between the *environmental context* targeted in the assessment and the physical location of the assessment; (2) possible confusion between *environmental context* and the *task type* categories in Aspect IV (Form); (3) participants focussing on one element of an assessment rather than selecting the category that best fits overall; and (4) the high level of information processing required by Delphi participants when applying definitions to case studies.

Aspect IV (Form)

Aspect IV has four components. These include: (1) a component that distinguishes between *standardised* or *non-standardised* administration procedures; (2) a component that describes the type of data collected in terms of *norm-referenced*, *criterion referenced* or *descriptive*; (3) a component that identifies assessments as either *static* or *dynamic*; and (4) a component that

describes *task type* in terms of discrete skill tasks versus contextualised or performance-based tasks and the naturalness of the communication interaction during assessment tasks. Each category in Aspect IV is mutually exclusive from other categories within each of the respective components (i.e. a particular assessment is either *standardised* or *non-standardised*; either *norm-referenced*, *criterion-referenced* or *descriptive*; either *static* or *dynamic* and is one *task type*).

At the end of round three, consensus on case study one was achieved with regards to all Aspect IV components, however case study two lacked consensus.. Participants comments reflected the following explanations for lack of consensus: (1) SLPs selected all possible ways an assessment could be used, rather than considering only the purposes for which assessments were used in cases studies; (2) it may be difficult to distinguish between the *task type* categories, *contextualised* and *activity-focussed*; (3) participants may ascribe definitions that are different to the taxonomy definitions when applying assessments in the respective case studies; and (4) the high level of information processing required from Delphi participants when applying the taxonomy definitions to case studies.

Discussion

In this study, a taxonomy for describing language assessments was developed, with experienced SLPs from a variety of work sectors reaching consensus on categories and definitions for describing the key features of assessments. Given the numerous challenges associated with the development of agreed-upon terminology, including the wide array of activities that may be undertaken when assessing the language abilities of children and the varied work sectors that span paediatric SLP practice (Walsh, 2005); this taxonomy represents a significant advancement in the field of child language assessment.

The use of case studies in the study allowed the application of the taxonomy to be examined and, in doing so, facilitated the refinement of the terms and definitions within the taxonomy. Nonetheless consensus across all components of the taxonomy with regards to categorisation of assessments was not reached for either case study. Greater inconsistency existed for case study two, particularly with regards to Aspect IV. Case study two described a language sampling procedure that followed a standardised procedure, but was dynamic in nature and provided descriptive data. It was noted in round one that assessment tools that are less prescriptive and more variable in terms of how they might be used, were less likely to be categorised consistently. The resource used in case study two was the Language Sampling Protocol (Westerveld & Gillon, 2002), a tool which may potentially have wide and varied applications by SLPs. It is possible that, despite all having the same case study, participants were drawn to considering how they themselves use the assessment tool, rather than how the tool's use was described in the case study. This may have contributed to this case study being less consistently categorised.

The components of the taxonomy that were not categorised consistently in case studies may also represent areas of SLP theoretical understanding that may need further development within the profession. For example, a lack of agreement on whether the assessments in the case studies targeted *semantics* and to a lesser extent, *executive functioning* persisted after round three. This lack of agreement may go beyond terminology and could reflect differences in professional understanding with regards to how these domains are assessed.

Environmental context also lacked consensus in both case studies, despite attempts to clarify this across Delphi rounds. While SLP literature discusses the value of assessing skills in

everyday communication environments, this discussion often occurs in the context of specific examples using terminology such as authentic or curriculum-based (Parsons et al., 2005; Schraeder et al., 1999). Similarly, while the concept of dynamic assessment is discussed across literature (Dockrell & Marshall, 2015), it is often presented as an assessment approach for children learning English as a second language and thus may not be an approach that SLPs in general paediatric language practice frequently identify themselves as using (Caesar & Kohler, 2009). This creates the possibility that, while SLPs agreed with the taxonomy distinctions for *environmental context* and *dynamic* assessment; applying the taxonomy may require SLPs to make more explicit and specific distinctions between assessments than they have previously been accustomed to making.

The identification of purposes for which assessments are used also emerged as an area of inconsistency in case studies, with participants tending to select many purposes for a single assessment. While it is important that assessment data be used maximally, it is also important that assessments are used for the purposes for which they have been designed (Newton, 2007). Researchers and clinicians must carefully decide which psychometric properties are most essential for a particular purpose and, thus, are most important to focus upon when selecting an assessment for that purpose (Wade, 2004). For example, assessments used for *diagnostic* and *screening* purposes should have established sensitivity and specificity data, whilst assessments used for *detecting change* should have evidence of responsiveness (Wade, 2004). While the extant SLP literature has focussed on assessments suited for *diagnostic* and *screening* purposes (Dockrell & Marshall, 2015), there appears to be limited literature explicitly examining assessments for purposes other than these, with most of the literature on assessment purpose originating from literature outside the SLP discipline (Newton, 2007; Wade, 2004). In the future, greater attention may need to be placed on the purposes of

different language assessments for SLP professional knowledge of language assessment practice to develop. Limitations with regards to current availability of assessments with established psychometric properties, particularly assessments that target communication from a more functional perspective (McLeod & Threats, 2008); may also lead to the use of assessments that are not ideally suited to the clinical purpose. It is also possible that constraints such as the high cost of commercial assessments, limited time to conduct assessments, or service provider policy demands may also overshadow decisions regarding the purposes for which assessment data is to be used (Fulcher-Rood, Castilla-Earls, & Higginbotham, 2018). Further investigation of assessment practices is necessary, particularly with regards to factors that influence SLP choice of assessment.

The findings of this Delphi study show that, even when consensus was reached on the categories and definitions within the taxonomy, consistent application of terminology cannot be assumed. The field of implementation science acknowledges challenges with knowledge to action transfer and the successful adoption of practice innovations (Miao, Power, & O'Halloran, 2015; Wilson, Brady, Lesesne, & NCCDPHP Work Group on Translation, 2011). These same challenges may apply to the adoption of new terminology. Although use of the taxonomy does not involve change to clinical practice per se, it may require SLPs to use new terminology or define terms related to assessment differently to what they may be accustomed to. Some terms may be engrained in particular organisations, service providers or in the minds of individual SLPs. In those circumstances, SLPs may need to develop an explicit understanding of how terminology in the taxonomy relates to the terminology they currently use in order to effectively “code switch” between terms. With this in mind, further research is needed to identify specific actions and strategies to assist consistent application of the taxonomy by SLPs when describing clinical practice (Wilson et al., 2011).

Limitations

Participants in this study represented a variety of geographical locations, work sectors and levels of professional experience, however, as with any Delphi study, it cannot be assumed that the same findings would be reached with a different group of participants. This study was also limited to Australia. Given that almost all of the background literature and research associated with the taxonomy originates from the United States or United Kingdom, it is expected that the terms in the taxonomy would also be applicable to other English speaking countries; however further research is warranted.

Participant drop-out over rounds poses a limitation in Delphi Studies (Boulkedid et al., 2011). In this study, agreement with the structure and definitions of the taxonomy was reached in round one, with 55 participants. Completion rates for round two and three were 74.4% and 78.2% respectively. Whilst this completion rate is reported as being typical in web-based surveys (Schleyer & Forrest, 2000), it may pose a limitation for *Part B* of the study in which categorisation of assessments using the taxonomy was examined; as it cannot be presumed that drop-out was random. Further, there was a large amount of reading required from participants in completing the surveys in this study, particularly in *Part B* which required reading the survey questions and taxonomy definitions alongside the case studies. While all attempts were made to present information in a reader-friendly manner, it is possible that categorisation of case studies was influenced by participant fatigue associated with high cognitive demand.

While the use of case studies served a purpose of allowing application of the taxonomy to be examined, the case studies are not without limitations themselves. It is possible that the use of

case studies may have drawn participants to considering the case (i.e. describing the domains that may require assessment based on the child's needs), rather than describing the specific assessment used in the case study. It was also not possible to comprehensively examine all aspects of the taxonomy using two case studies. Results may have been different if the case studies used other types of assessments.

Conclusion

In this study, a taxonomy for describing child language assessment practices was developed and a rigorous methodology applied in order to evaluate the consensus of it amongst a group of experienced paediatric SLPs. The high level of agreement from clinicians and academics with the taxonomy structure and definitions represents a significant step towards addressing some of the challenges that inconsistent terminology poses for the field of child language assessment. The taxonomy provides structure, terminology and definitions from which further professional knowledge and future research may be built upon (Eadie, 2003). It has uses for the collection of data on SLP assessment practices, provision of SLP training, and for making comparisons between different assessments in research studies. Given that some components of the taxonomy were not consistently applied when describing the case studies, further research is recommended to identify strategies that support implementation of the taxonomy.

Acknowledgments: The authors would like to acknowledge that an Australian Government Research Training Program Scholarship supported this research.

References

- American Speech and Hearing Association. (2018). Assessment tools, techniques, and data sources Retrieved from <https://www.asha.org/Practice-Portal/Clinical-Topics/Late-Language-Emergence/Assessment-Tools-Techniques-and-Data-Sources/>
- Barnes, S., & Bloch, S. (2018). Why is measuring communication difficult? A critical review of current speech pathology concepts and measures. *Clinical Linguistics & Phonetics*. doi:<https://dx.doi.org/10.1080/02699206.2018.1498541>
- Bishop, D. V. M. (2003). *Children's Communication Checklist - Second Edition (CCC-2)*: Pearson.
- Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review. *PLoS ONE*, 6(6), e20476. doi:10.1371/journal.pone.0020476
- Caesar, L. G., & Kohler, P. D. (2009). Tools clinicians use: A survey of language assessment procedures used by school-based speech-pathologists. *Communication Disorders Quarterly*, 30(4), 226-236. doi:10.1177/1525740108326334
- Cowie, M., Wanger, K. M., Cartwright, A., Bailey, H., Millar, J. A., & Price, M. (2001). A review of Clinical Terms Version 3 (Read Codes) for speech and language record keeping. *International Journal of Language & Communication Disorders*, 36(1), 117-126. doi:10.1080/13682820116848
- DeJong, G., Horn, S. D., Gassaway, J. A., Slavin, M. D., & Dijkers, M. P. (2004). Toward a taxonomy of rehabilitation interventions: Using an inductive approach to examine the “black box” of rehabilitation. *Archives of Physical Medicine and Rehabilitation*, 85(4), 678-686. doi:10.1016/j.apmr.2003.06.033

- Dewart, H., & Summers, S. (1995). Pragmatics Profile of Everyday Communication Skills in Children. Retrieved from http://complexneeds.org.uk/modules/Module-2.4-Assessment-monitoring-and-evaluation/All/downloads/m08p080c/the_pragmatics_profile.pdf
- Dockrell, J. E., & Marshall, C. R. (2015). Measurement issues: Assessing language skills in young children. *Child and Adolescent Mental Health*, 20(2). doi:10.1111/camh.12072
- Eadie, P. (2003). Speech pathology assessment practices: One assessment or many? *Advances in Speech Language Pathology*, 5(1), 65-68. doi:10.1080/14417040510001669081
- Fulcher-Rood, K., Castilla-Earls, A. P., & Higginbotham, J. (2018). School-based speech-language pathologists' perspectives on diagnostic decision making. *American Journal of Speech-Language Pathology*, 27(2), 796-812. doi:10.1044/2018_AJSLP-16-0121
- Hegde, M. N., & Pomaville, F. (2017). Assessment of communication disorders in children. In *Assessment of communication disorders in children*. US: Plural Publishing.
- Hsiu-Fang, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 16(9), 1277-1288. doi:10.1177/104973230527668
- Hughes, J., & Orange, J. B. (2007). Mapping functional communication measurements for traumatic brain injury to the WHO-ICF. *Revue canadienne d'orthophonie et d'audiologie*, 31(3), 135-142.
- IBM Corp. (Released 2011). IBM SPSS Statistics for Windows Version 20.0. Armonk, NY: IBM Corp.
- Kaderavek, J. M. (2015). Assessment of language disorders. In *Language disorders in children: Fundamental concepts of assessment and intervention*. US: Pearson.
- Lahey, M. (1988). *Language disorders and language development*. New York: MacMillan.

- Larson, V. L., & McKinley, N. L. (2007). Procedures for direct assessment. In *Communication solutions for older students* (pp. 209-268). US: Thinking Publications.
- McLeod, S., & Threats, T. T. (2008). The ICF-CY and children with communication disabilities. *International Journal of Speech-Language Pathology*. doi:10.1080/1754950070183469
- Miao, M., Power, E., & O'Halloran, R. (2015). Factors affecting speech pathologists' implementation of stroke management guidelines: A thematic analysis. *Disability and Rehabilitation*, 37(8), 674-685. doi:10.3109/09638288.2014.932444
- Newton, P. E. (2007). Clarifying the purposes of educational assessment *Assessment in Education: Principles, Policy & Practice*, 14(2), 149-170. doi:10.1080/09695940701478321
- Parsons, S., Law, J., & Gascoigne, M. (2005). Teaching receptive vocabulary to children with Specific language Impairment: A curriculum-based approach. *Child Language Teaching and Therapy*, 21(1), 39-59. doi:10.1191/0265659005ct280oa
- Paul, R., & Norbury, C. F. (2012). Chapter 2: Assessment. In R. Paul & C. F. Norbury (Eds.), *Language disorders from infancy through adolescence: Listening, speaking, reading, writing and communicating* (4th ed., pp. 22-60). Canada: Mosby Elsevier.
- Roulstone, S. (2001). Consensus and variation between speech and language therapists in the assessment and selection of preschool children for intervention: A body of knowledge or idiosyncratic decisions? *International Journal of Language and Communication Disorders*, 36(3), 329–348. doi:10.1080/1368282001001992 8
- Roulstone, S. (2015). Exploring the relationship between client perspectives, clinical expertise and research evidence. *International Journal of SpeechLanguage Pathology*, 17(3), 211-222. doi:10.3109/17549507.2015.1016112

- Schleyer, T. K., & Forrest, J. L. (2000). Methods for the design and administration of web-based surveys. *Journal of the American Medical Informatics Association*, 7(4), 416-425. doi:10.1136/jamia.2000.0070416
- Schraeder, T., Quinn, M., Stockman, I. J., & Miller, J. (1999). Authentic assessment as an approach to preschool speech-language screening. *American Journal of Speech-Language Pathology*, 8(3), 195-200. doi:10.1044/1058-0360.0803.195
- Tapio, P., Paloniemi, R., Varho, V., & Vinnari, M. (2011). The unholy marriage? Integrating qualitative and quantitative information in Delphi processes *Technological Forecasting & Social Change*, 78, 1616 –1628. doi:10.1016/j.techfore.2011.03.016
- Wade, D. T. (2004). Assessment, measurement and data collection tools. *Clinical Rehabilitation*, 18, 233-237. doi:10.1191/0269215504cr183ed
- Walsh, R. (2005). Meaning and purpose: A conceptual model for speech pathology terminology. *Advances in Speech-Language Pathology*, 7(2), 65-76. doi:10.1080/14417040500125285
- Walsh, R., & IGOTF-CSD. (2006). A history of terminology: International group on terminology frameworks – communication science and disorders Retrieved from http://www.dhrs.uct.ac.za/sites/default/files/image_tool/images/147/History_of_CSD.pdf
- Westby, C., & Washington, K. N. (2017). Using the International Classification of Functioning, Disability and Health in assessment and intervention of school-aged children with language impairments. *Language, Speech, and Hearing Services in Schools*, 48, 137-152. doi:10.1044/2017_LSHSS-16-0037
- Westerveld, M., & Claessen, M. (2014). Clinician survey of language sampling practices in Australia. *International Journal of Speech-Language Pathology*, 16(3), 242-249. doi:10.3109/17549507.2013.871336

Westerveld, M., & Gillon, G. T. (2002). Language Sampling Protocol. Retrieved from <http://www.education.canterbury.ac.nz/documents/gillon/languageprotocol.pdf>

Wiig, E. H., Semel, E., & Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals - 4th Edition (CELF-4)* (5 ed.): Pearson.

Wilson, K. M., Brady, T. J., Lesesne, C., & NCCDPHP Work Group on Translation. (2011). An organizing framework for translation in public health: The knowledge to action framework. *Preventing Chronic Disease*, 8(2), 1-7.

World Health Organisation. (2015). *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (Vol. 10). Geneva, Switzerland: World Health Organisation.

Table I. Participant demographics

Category	Subcategory	Round One n (%)	Round Two n (%)	Round Three n (%)
State	QLD	7 (12.7%)	7 (16.3%)	5 (15.6%)
	NSW	10 (18.2%)	7 (16.3%)	6 (18.8%)
	ACT	1 (1.8%)	1 (2.3%)	0 (0.0%)
	VIC	16 (29.1%)	11 (25.6%)	9 (28.1%)
	TAS	3 (5.5%)	2 (4.7%)	2 (6.3%)
	NT	3 (5.5%)	1 (2.3%)	1 (3.1%)
	SA	7 (12.7%)	7 (16.3%)	6 (18.8%)
	WA	8 (14.5%)	7 (16.3%)	3 (9.4%)
	Total	55	43	32
Current Employment*	Health Sector (government or non-government)	5 (9.1%)	5 (11.6%)	2 (6.3%)
	Education Sector (government or non-government)	18 (32.7%)	17 (39.5%)	16 (50.0%)
	Private Practice/Small Business	10 (18.2%)	7 (16.3%)	4 (12.5%)
	University	13 (23.6%)	10 (23.3%)	7 (21.9%)
	Other agency (government or non-government)	3 (5.5%)	2 (4.7%)	0 (0.0%)
	Currently not working as SLP	1 (1.8%)	0 (0.0%)	0 (0.0%)
	Work across two of the above sectors	5 (9.1%)	2 (4.7%)	3 (9.4%)
	Total	55	43	32
	Qualifications in addition to Bachelor or Graduate Equivalent degree*	Masters or PhD	24 (43.6%)	18 (41.9%)
Diploma (Education or Psychology)		2 (3.6%)	2 (4.7%)	1 (3.1%)
No other qualifications		29 (52.7%)	23 (53.5%)	16 (50.0%)
Total		55	43	32
Years of experience (Full-time equivalent) *	5-10 years	10 (18.2%)	7 (16.3%)	5 (15.6%)
	11-15 years	10 (18.2%)	9 (20.9%)	8 (25.5%)
	16-21 years	13 (23.6%)	9 (20.9%)	9 (28.1%)
	21-30 years	12 (21.8%)	9 (20.9%)	5 (15.6%)
	30+ years	10 (18.2%)	9 (20.9%)	5 (15.6%)
	Total	55	43	32

Note: *As reported by participant

Table II. Agreement with structure of taxonomy and definitions (*Part A*)

Aspect of the taxonomy	Results					
	Median		IQR		Percentage agreement	
	Round 1 n=55	Round 2 n= 42	Round 1 n=55	Round 2 n=42	Round 1 n=55	Round 2 n=42
Aspect I Structure	4	5	1	1	92.7 (51)	97.7 (41)
Aspect I Definitions	4.5	4	1	1	90.9 (50)	93.0 (39)
Aspect II Structure	4	5	1	1	92.7 (51)	95.2 (40)
Aspect II Definitions	4	5	1	1	96.4 (53)	93.0 (39)
Aspect III Structure	4	5	1	1	87.3 (48)	90.7 (38)
Aspect III Definitions	4.5	5	1	1	90.9 (50)	88.4 (37)
Aspect IV Structure	4.5	5	1	1	96.4 (53)	97.7 (41)
Aspect IV Definitions	4.5	5	1	1	98.2 (54)	95.2 (40)
Overall Structure ^a	4	5	1	1	100 (55)	100 (42)

Table Key:

Percentage agreement: Percentage of participants who selected “agree” or “strongly agree”

Scale: 5 = Strongly Agree, 4 = Agree, 3 = Neither Agree or Disagree, 2 = Disagree, 1 = Strongly Disagree

Median: The value that appears most often (i.e., the most frequently selected answer)

IQR: Inter-quartile Range i.e. the middle 50% of the data (i.e. the difference between 75th and 25th percentiles)

^aDuring round one, 54 participants completed this question

Table III. Agreement with categorisation of assessments in case studies (*Part B*)

Aspect within taxonomy	Categories within aspect	Results			
		Case Study 1		Case Study 2	
		Round 2 n=43 % of participants who selected category	Round 3 n= 32 % of participants who selected category	Round 2 n=43 % of participants who selected category	Round 3 n=32 % of participants who selected category
Aspect I Categories not mutually exclusive In round three participants could only choose one category in addition the categories already agreed-upon in round two.	Spoken	97.7 ^a	NA	100 ^a	NA
	Written	2.3	NA	2.3	NA
	Semantics	62.8^b	37.5^b	76.7^c	56.3^b
	Morphosyntax	7.0	NA	86.0 ^a	NA
	Social Abilities	100 ^a	NA	37.2	NA
	Discourse	18.6	NA	97.7 ^a	NA
	Meta Abilities	7.0	NA	18.6	NA
	Executive Functions	30.2^b	28.1^b	25.6^b	25.0
	Comprehension	83.7 ^a	NA	100 ^a	NA
	Production	100 ^a	NA	100 ^a	NA
Aspect II Categories not mutually exclusive. In round three participants could only choose one other prognostic and one other analytical category in addition the categories already agreed-upon in round two.	Predict Outcome	25.6^b	25.0	58.1^{ab}	34.4^{ab}
	Select Intervention	79.1 ^a	NA	72.1^b	43.8^b
	Plan Dosage	39.5^b	25.0	41.9^b	NA
	Screening	30.2^b	31.3^b	20.9	NA
	Diagnostic	41.9^b	31.3^b	46.5^b	15.6
	Detect Change	23.3	NA	37.2^{ab}	78.1 ^a
	Describe Status	87.7 ^a	NA	88.4 ^a	NA
Aspect III Categories mutually exclusive	SLP Conducted	39.5^b	15.6	95.3 ^a	NA
	Other Conducted	0.0	NA	2.3	NA
	Software	0.0	NA	0.0	NA
	Proxy - Reported	60.^{ab}	84.4 ^a	2.3	NA
	Clinic	34.9^b	28.1^b	53.5^{ab}	53.1^{ab}
Categories mutually exclusive	Community - Home	58.1^{ab}	71.9^{ab}	0.0	NA
	Community - School	0.0	NA	44.2^b	46.9^b
	Community - Other	7.0	NA	2.3	NA
Aspect IV Categories mutually exclusive	Standardised	20.9	NA	30.2^{ab}	56.3^{ab}
	Non-standardised	79.1 ^a	NA	69.8^b	43.8^b
Categories mutually exclusive	Norm-referenced	0.0	NA	7.0	NA
	Criterion-referenced	11.6	NA	14.0	NA
Categories mutually exclusive	Descriptive data	88.4 ^a	NA	79.1 ^a	NA
	Static	86.0 ^a	NA	39.5^b	43.8^b
	Dynamic	14.0	NA	60.5^{ab}	56.3^{ab}
Categories mutually exclusive	Hierarchical	9.3	9.4	4.7	15.6
	Non-Hierarchical	14.0	6.3	16.3	25.0
	Contextualised	65.1^{ab}	78.1 ^a	48.8^{ab}	56.3^{ab}
	Activity-focussed	11.6	6.3	30.2^b	3.1

^a = categories researchers expected would be selected for each case study

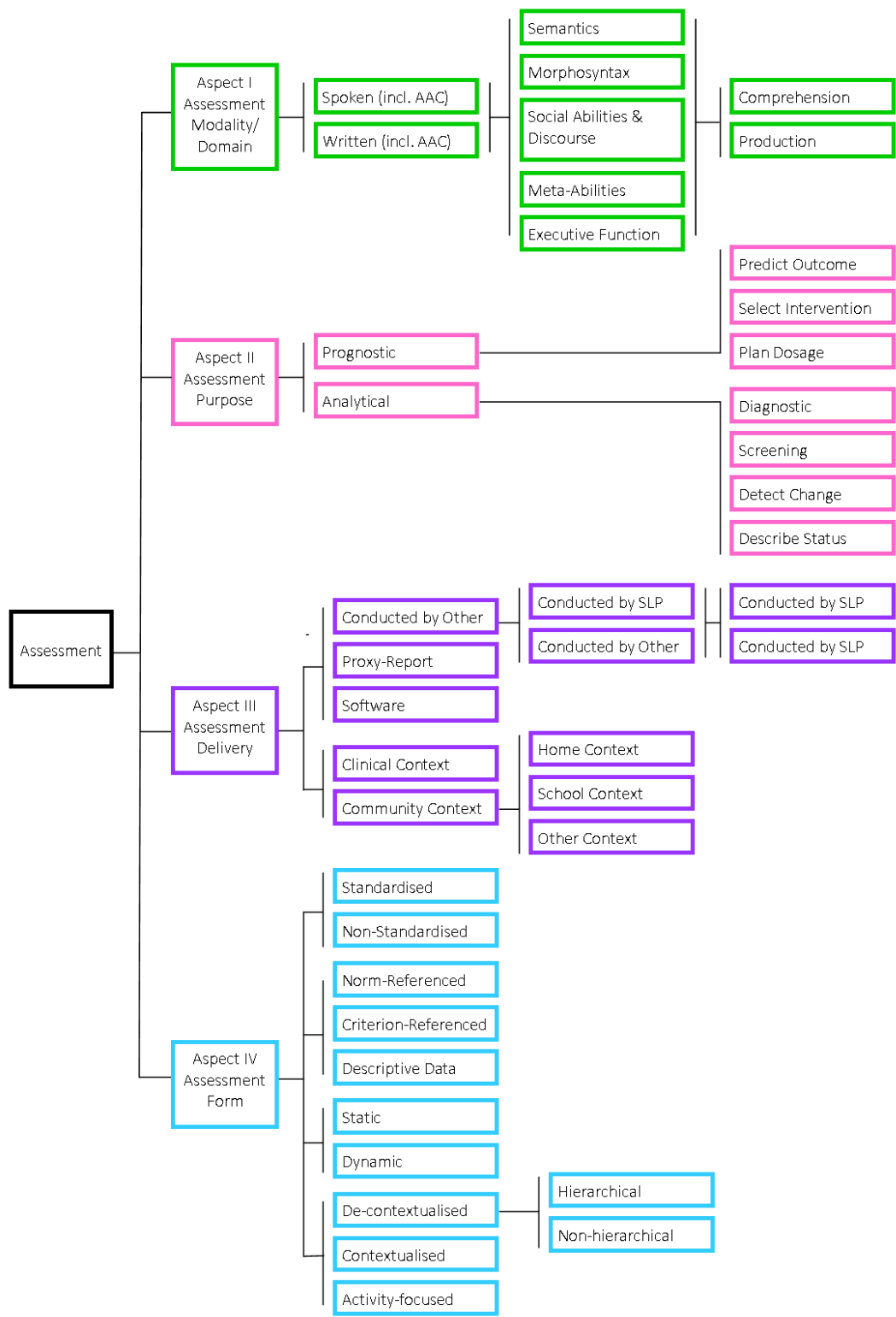
^b = categories where inconsistency was identified i.e., 25-75% of participants selected this category

^c = categories where inconsistency was identified due to an unexpected result i.e. this category reached criteria for consensus, however consensus did not align with researcher expectations

NA = this question was not asked in round three as consensus was reached in round two

Case Study 1: Parent interview using Pragmatics Profile

Case Study 2: Language sampling using Language Sampling Protocol



**Describing language assessments for school-aged children: A Delphi study
Supplementary Table I.**

Definitions of terminology within the assessment taxonomy

ASPECT I (Modalities and Domains)	
Term and definition	Examples of assessments
<p>Spoken Language: Language exchanged verbally, or via an alternative in situations where peers would typically use verbal communication (includes pre-linguistic communication)</p> <p>(American Speech and Hearing Association, 1993; Beukelman & Mirenda, 2013).</p>	<ul style="list-style-type: none"> • <i>Assessment of spoken communication via a single mode (single-modality) e.g. Speech-only or AUSLAN</i> • <i>Assessment of spoken communication via multiple modes (multi-modal) e.g. Key-word sign or Aided language stimulation</i>
<p>Written Language: Language exchanged through text (print) or via an alternative in situations where peers would be typically be reading or writing</p> <p>(American Speech and Hearing Association, 1993; Beukelman & Mirenda, 2013).</p>	<ul style="list-style-type: none"> • <i>Assessment of written communication via a single mode (single-modality) e.g. Text-only</i> • <i>Assessment of written communication via multiple modes (multi-modal) e.g. Text with symbol support</i>
<p>Semantics: Understanding and expression of words and word meanings (e.g. vocabulary, word retrieval, lexical meaning).</p> <p>(American Speech and Hearing Association, 1993; Apel, 2014; Beukelman & Mirenda, 2013; Dockrell & Marshall, 2015; Paul & Norbury, 2012).</p>	<ul style="list-style-type: none"> • <i>Knowledge of vocabulary words is assessed by having the child name a series of pictures</i> • <i>A sample of a child's language is analysed for number of different words (NWD) or type-token ratio (TTR)</i> • <i>Semantic knowledge is assessed by asking the child to give synonyms and antonyms for different words</i>
<p>Morphosyntax: Understanding and expression of different word forms and the order and combination of words in sentences</p> <p>(American Speech and Hearing Association, 1993; Apel, 2014; Beukelman & Mirenda, 2013; Dockrell & Marshall, 2015; Paul & Norbury, 2012).</p>	<ul style="list-style-type: none"> • <i>Sentence structure is assessed by asking a child to point to pictures that represent a spoken sentence</i> • <i>A sample of a child's language is analysed for MLU and Brown's Grammatical Morphemes</i>

<p>Social Abilities and Discourse (Pragmatics): Giving and making meaning in social context or communication for social purposes. Includes:</p> <ul style="list-style-type: none"> - Pre-linguistic communication e.g. facial expression, joint attention, gesturing etc - Communication intentions/purposes e.g. Requesting, commenting, greetings, asking questions, giving reasons, making predictions etc - Non-verbal communication e.g. understanding emotions from body language and facial expressions - Non-literal language e.g. inferences, idioms, metaphors, jokes, sarcasm etc - Matching communication style to social context e.g. Adjusting communication style between friends and teachers - Conversation conventions e.g. topic selection, topic maintenance, conversational turn-taking etc - Text cohesion e.g. verbal fluency (mazes and incomplete sentences), transitions between sentences/paragraphs etc - Text organisation (discourse or macrostructure) e.g. Narrative structure (story grammar), episodic structure etc <p>(American Speech and Hearing Association, 1993; Apel, 2014; Beukelman & Mirenda, 2013; Dockrell & Marshall, 2015; Paul & Norbury, 2012).</p>	<ul style="list-style-type: none"> • Information on the range of communication functions for which a child communicates is profiled during a parent interview • Narrative structure (story grammar) and text cohesion are assessed during a narrative retell task • Non-verbal communication and conversation conventions are observed during a conversation between the child and the SLP
<p>Meta-Abilities: Ability to think about own thought processes and understand how to regulate these processes for effective learning. Includes:</p> <ul style="list-style-type: none"> - Meta-cognition: Knowledge and use of strategies for managing and self-monitoring own learning. - Meta-Language: Knowledge of phonemic (phonemic awareness), morphological/syntactic (meta-syntactic) or text-level (meta-narrative) rules in relation to own skills; and ability to effectively apply these rules for improved performance. - Meta-pragmatics: Knowledge of social conventions in relation to own communication and ability to apply this knowledge to improve communication with others <p>(Kamhi, Masterson, & Apel, 2007; Larson & McKinley, 2007; Law, Campbell, Roulstone, Adams, & Boyle, 2007; Starling, Munro, Togher, & Arciuli, 2012)</p>	<ul style="list-style-type: none"> • A child is asked to describe strategies that facilitate their own learning or performance (meta-cognition) • A child describes the features of a narrative story and their understanding of what constitutes good narrative structure (meta-language) • Phoneme awareness skills are assessed by asking the child to identify the number of phonemes in words (meta-language) • A child is asked to identify what they would do in a given social situation and why (meta-pragmatics)
<p>Executive Functions: Collection of related cognitive processes necessary for execution of goal-directed, controlled, purposeful behaviour. Includes:</p> <ul style="list-style-type: none"> - Inhibition (self-control): Ability to focus and attend to tasks through suppression of inappropriate thoughts, comments and behaviours - Emotion control (self-regulation): Ability to manage emotions for goal achievement and task completion - Working memory: Ability to retain, process and manipulate pieces of information for short periods of time to complete required tasks - Organisation: (strategic planning) Ability to use organisational strategies for task completion e.g. envisioning the end product, planning steps to complete tasks, identifying solutions to problems etc - Mental flexibility: Ability to integrate prior knowledge and experiences or effectively apply of different rules for different situations - Sustained attention: Ability to maintain attention to tasks despite distractions or fatigue <p>(Hyter, 2003; Montgomery, Magimairaj, & Finney, 2010; Serry, Rose, & Liamputtong, 2008; Singer & Bashir, 1999; Ukrainetz, 2006)</p>	<ul style="list-style-type: none"> • Auditory working memory is assessed by asking the child to repeat strings of numbers or words • Organisational skills are assessed by observing a child in class while they plan out a project by setting goals and identifying steps involved. • Inhibitory control is examined through a task that requires the child to read names of colours written in coloured ink that does not match the word that's spelled out i.e. the child must say the colour they see, as opposed to the word that is written

<p>Comprehension: Understanding of information, knowledge and ideas communicated by others (includes verbal and non-verbal).</p> <p>(American Speech and Hearing Association, 1993)</p>	<ul style="list-style-type: none"> • <i>A child's ability to understanding and follow directions is assessed by asking the child to follow a series of instructions</i> • <i>A child's understanding of facial expressions is assessed by asking the child to point to faces that display different emotions</i>
<p>Production: Ability to convey information, knowledge and ideas to others (includes verbal or non-verbal).</p> <p>(American Speech and Hearing Association, 1993)</p>	<ul style="list-style-type: none"> • <i>A child's vocabulary is assessed in a picture naming task</i> • <i>A child's ability to produce a story is assessed in a narrative retell task</i>
<p>ASPECT II (Assessment Purpose)</p>	
<p>Term and definition</p>	<p>Examples</p>
<p>Predict outcome: Identify risk of poor future outcome, predict need for intervention or identify support needs.</p> <p>(Olswang & Bain, 1996; Vaz et al., 2015; Wade, 2004; Wixson & Valencia, 2011)</p>	<ul style="list-style-type: none"> • <i>Support needs at school (type/level of curriculum differentiation or special education support) are identified by assessing performance in the presence of different prompts or scaffolds (i.e. dynamic assessment using graded prompting).</i> • <i>Early primary school or kindergarten children are assessed on pre-literacy skills that are seen as predictive of later literacy success (to identify those who may benefit from participation in a preventive program)</i>
<p>Select intervention: Identify suitability for an intervention approach or select intervention targets.</p> <p>(Eadie, 2003; Newton, 2007; Paul & Norbury, 2012; Vaz et al., 2015; Wade, 2004; Westerveld & Claessen, 2014; Wixson & Valencia, 2011)</p>	<ul style="list-style-type: none"> • <i>An interview with parents (regarding family preferences/concerns, child's likes/dislikes, available resources etc) assists with selection an intervention approach.</i> • <i>A child's ability to produce a range of different morphological and syntactical forms is assessed to identify the forms to be targeted in intervention.</i>
<p>Plan dosage: Predict intensity (dosage) of intervention.</p> <p>(Vaz et al., 2015; Wade, 2004; Westby, 2007)</p>	<p><i>The amount of intervention needed to achieve an outcome is estimated by:</i></p> <ul style="list-style-type: none"> • <i>Assessing a child's response to a short trial of the intervention (dynamic assessment in a test-teach-retest format)</i> • <i>Collecting a comprehensive history regarding the child's response to previous interventions (response to intervention).</i>
<p>Screening: Identify children who may have a disorder that requires further diagnostic assessment to confirm.</p> <p>(American Speech-Language-Hearing Association, 2004; Dockrell & Marshall, 2015; Eadie, 2003; Paul & Norbury, 2012; Vaz et al., 2015; Wade, 2004; Westerveld & Claessen, 2014; Wixson & Valencia, 2011)</p>	<ul style="list-style-type: none"> • <i>Assessment is conducted to identify if diagnostic assessment should be conducted and/or the domains to be targeted in diagnostic assessment</i>

<p>Diagnostic: Diagnose a condition or make a comparison with peers.</p> <p>(American Speech-Language-Hearing Association, 2004; Betz, Eickhoff, & Sullivan, 2013; Dockrell & Marshall, 2015; Eadie, 2003; Kapantzoglou, Restrepo, & Thompson, 2012; Paul & Norbury, 2012; Vaz et al., 2015; Wade, 2004; Westerveld & Claessen, 2014; Wixson & Valencia, 2011)</p>	<ul style="list-style-type: none"> • <i>Assessment conducted to identify the presence or severity of a diagnosis; or determine if functioning is different to peers</i>
<p>Detect change: Measure change in status or monitor progress over time.</p> <p>(Eadie, 2003; Paul & Norbury, 2012; Vaz et al., 2015; Wade, 2004; Westerveld & Claessen, 2014; Wixson & Valencia, 2011)</p>	<ul style="list-style-type: none"> • <i>Assessment repeated at different intervals to monitor progress over time</i> • <i>Pre & post intervention assessment to document change (or no change in a control group)</i>
<p>Describe status: Assessment for the purpose of describing or explaining a particular aspect of a student's functioning.</p> <p>(Vaz et al., 2015; Wade, 2004; Wixson & Valencia, 2011)</p>	<ul style="list-style-type: none"> • <i>Communicative behaviours are described (gesture dictionary) in order to help unfamiliar communication partners understand/interpret a student's communication behaviours</i> • <i>An SLP assesses a student's performance on spoken comprehension tasks to further explore reasons why others report that the student has difficulties understanding verbal information, despite the student achieving an average score on a standardised receptive language test.</i>
<p>ASPECT III (Assessment Delivery)</p>	
<p>Term and definition</p>	<p>Examples</p>
<p>By Person - Conducted by SLP: Assessment is conducted by an SLP through pre-planned observation, testing or sampling of a child's skills. Results may be analysed at the time or may be analysed later from an audio/video recording. Others may assist with administration or technology may be used to score; however, the SLP has the primary role in planning the assessment and analysing findings.</p> <p>(Kaminski, Abbott, Aguayo, Latimer, & Good, 2014; Wixson & Valencia, 2011)</p>	<ul style="list-style-type: none"> • <i>An SLP conducts a standardised assessment</i> • <i>An SLP transcribes and analyses a language sample that was audio-recorded earlier by a teacher</i> • <i>An SLP compares and analyses a narrative transcript with reference to a database of normative data from peers</i>
<p>By Person - Conducted by Other: Assessment conducted by another person (teacher, another professional etc), through pre-planned observation, testing or sampling of the child's skills. An SLP may provide training or support to the other person, or technology may be used (e.g. online stimulus materials or software that calculates test scores); however, the other person has the primary role in planning the assessment and analysing/interpreting results.</p> <p>(Kaminski et al., 2014; Wixson & Valencia, 2011)</p>	<ul style="list-style-type: none"> • <i>A teacher assesses the phonemic awareness skills of a group of children with literacy difficulties to determine literacy intervention goals for those children.</i>
<p>Face-to-face (only for assessments conducted by a person): Assessment is conducted with the child and an assessor in the same room.</p> <p>(Edwards, Stredler-Brown, & Houston, 2012; Mashima & Doarn, 2009; Waite, Theodoros, Russell, & Cahill, 2010a, 2010b)</p>	<ul style="list-style-type: none"> • <i>During a face-to-face interaction with a child, an SLP audio-records a language sample for later analysis</i> • <i>An SLP administers a standardised test face-to-face and scores with the assistance of scoring software</i>

<p>ICT (only for assessments conducted by a person): Assessments is conducted with the assessor and the child communicating through ICTs (information and communication technologies). Technology that is not used for two-way communication between individuals during the assessment is not considered ICT (e.g. audio/video recorders)</p> <p>(Edwards et al., 2012; Mashima & Doarn, 2009; Waite et al., 2010a, 2010b)</p>	<p><i>Assessments conducted by:</i></p> <ul style="list-style-type: none"> • <i>Web-conferencing (such as Skype or Zoom)</i> • <i>Video-conferencing</i> • <i>Telephone</i>
<p>Proxy-Report: Skills are not assessed in the moment they occur, but are documented based on retrospective reports from others, such as in an interview or by completion of questionnaire/checklist. The reported information:</p> <ul style="list-style-type: none"> - may be from a child (self-report), another professional, a caregiver, a teacher or a peer. - may relate to previous skills (e.g. developmental or educational history) or current abilities (e.g. current level of development; or performance in the current unit of schoolwork). <p>(Bishop & McDonald, 2009; Dockrell & Marshall, 2015; Schraeder, 2008; Williams, 2006)</p>	<ul style="list-style-type: none"> • <i>During a case history interview, a parent reports on information about a child's history that may be diagnostically significant.</i> • <i>A teacher reports information by completing a checklist regarding the pre-linguistic behaviours they have observed the child use at school.</i>
<p>Software delivered: The child's abilities are assessed through a predominantly computerised procedure with no (or extremely little) input from a person. Software program selects/presents tasks, records data and scores results. A person may set a child up at a computer or be present to supervise while the child sits the test. If a person is required to administer items, respond to the child's test answers, record observations or score results; then the assessment is not categorised as software.</p> <p>(Ockey, 2009; Richards et al., 2017)</p>	<p><i>Assessments conducted by:</i></p> <ul style="list-style-type: none"> • <i>App or web-based program</i> • <i>Computer (software) program</i>
<p>Clinical context: Skills are assessed within a clinical context i.e. the assessment does not incorporate materials or communication partners from the day-to-day environment. Note: This category refers to the context being assessed; not physical location</p> <p>(Schraeder, Quinn, Stockman, & Miller, 1999; Westby, 2007)</p>	<ul style="list-style-type: none"> • <i>A child is withdrawn from regular classroom activities for narrative assessment by an SLP using materials that the SLP has brought to the school. Although the child is at school, the assessment context is that of a clinical environment</i> • <i>An SLP administers a standardised assessment at the child's home in a quiet room away from distractions. The assessment is conducted according to administration guidelines and does not incorporate any of the activities, materials or people that the child interacts with at home.</i>
<p>School context: Communication is assessed in a school (or Kindergarten) context i.e. uses communication partners, communication situations or materials that represent a school environment. Note: This category refers to the context being assessed; not physical location</p> <p>(Nelson, 1989; Parsons, Law, & Gascoigne, 2005; Schraeder, 2008; Schraeder et al., 1999; Westby, 2007)</p>	<ul style="list-style-type: none"> • <i>In an interview with the SLP, a teacher is asked to comment on how the child communicates with teachers and classmates during whole class lessons</i> • <i>An SLP assesses a child's oral and reading comprehension skills using the text being studied in the current unit of English and activities similar to those used to teach the English school curriculum</i>
<p>Home context: Communication is assessed in a home context i.e. uses communication partners, communication situations or materials that represent a home environment. Note: This category refers to the context being assessed; not physical location</p> <p>(Schraeder, 2008; Schraeder et al., 1999; Westby, 2007)</p>	<ul style="list-style-type: none"> • <i>During an appointment in an outpatient clinic, a parent completes a checklist based on the communication behaviours they have observed at home</i> • <i>An SLP observes a child play and read with his mother and siblings using similar toys and books as those in the child's home. Although the child is in a clinic consultation room, the assessment context is considered to be representative of a home environment</i>

<p>Other community context: Communication is assessed in a community context i.e. uses communication partners, communication situations or materials that represent a community environment. Note: This category refers to the context being assessed; not physical location</p> <p>(Schraeder, 2008; Schraeder et al., 1999; Westby, 2007)</p>	<ul style="list-style-type: none"> • <i>A child describes the communication difficulties they experience when interacting coaches and teammates during extra-curricular soccer training and matches</i> • <i>A child's skills are observed and recorded during a work experience placement (e.g. interacting with customers, taking orders and counting money)</i>
<p>ASPECT IV (Assessment Form)</p>	
<p>Term and definition</p>	<p>Examples</p>
<p>Standardised: Assessments designed to be administered and scored in a consistent manner, which is the same for all children who are assessed i.e. specific questions or tasks, clear administration and scoring guidelines, defined assessment materials and set procedures to elicit responses from the child.</p> <p>(Hegde & Pomaville, 2017; Paul & Norbury, 2012)</p>	<ul style="list-style-type: none"> • <i>Use of a language sampling that follows specific administration procedures, including use of set materials and specific prompts to elicit the retell from the child</i>
<p>Non-standardised: Assessments that may not be administered the same way by different assessors in different conditions. Procedures for administration and scoring may be variable or may not be described well enough for consistent administration and scoring.</p> <p>(Hegde & Pomaville, 2017; Paul & Norbury, 2012)</p>	<ul style="list-style-type: none"> • <i>Use of a language sampling procedure that does not have set administration guidelines i.e. a task that the SLP has created themselves or adapted from another resource.</i>
<p>Norm-referenced: Assessments that quantitatively compare a child's score to scores from a sample of matched peers who completed the same task. These assessments should always be standardised.</p> <p>(Caesar & Kohler, 2009; Paul & Norbury, 2012; Schraeder, 2008; Ukrainetz, 2015a)</p>	<ul style="list-style-type: none"> • <i>A child's performance is compared to normative scores (standard scores means or percentile ranks) derived from a sample of similar peers</i>
<p>Criterion-referenced: Assessments that compare a child's performance against a pre-determined level or criterion (i.e. skills expected given a child's age, grade or curriculum level). These assessments may or may not be standardised.</p> <p>(Caesar & Kohler, 2009; Paul & Norbury, 2012; Schraeder, 2008; Ukrainetz, 2015a)</p>	<ul style="list-style-type: none"> • <i>A child's performance is compared to the curriculum expectations for their year level</i> • <i>A child's syntactical and morphological are assessed in relation to knowledge of developmental expectations</i>
<p>Descriptive: Assessments designed to give descriptive or qualitative data on a child's abilities. These assessments may or may not be standardised.</p> <p>(Caesar & Kohler, 2009; Paul & Norbury, 2012; Schraeder, 2008; Ukrainetz, 2015a)</p>	<ul style="list-style-type: none"> • <i>A child's narrative retell skills are described in terms of strengths and weaknesses</i> • <i>A child's social abilities are described in terms of functional abilities observed in the classroom</i>

<p>Static: Assessment procedures that are designed to measure performance at a given point in time under given conditions.</p> <p>(American Speech-Language-Hearing Association, 2004; Caesar & Kohler, 2009; Dockrell & Marshall, 2015; Kapantzoglou et al., 2012; Leaders Project, 2013; Peña et al., 2006)</p>	<ul style="list-style-type: none"> • A child's vocabulary knowledge is assessed in a picture naming task that compares performance to peers of the same age
<p>Dynamic: Assessment procedures designed to assess a child's performance under varied conditions or investigate response to intervention. These describe learning potential or identify successful supports and teaching techniques. Includes:</p> <ul style="list-style-type: none"> - Test-teach-retest procedures - Testing the limits (response to task modification) - Graded levels of prompting (response to different levels of prompting) <p>(American Speech-Language-Hearing Association, 2004; Caesar & Kohler, 2009; Dockrell & Marshall, 2015; Kapantzoglou et al., 2012; Leaders Project, 2013; Peña et al., 2006)</p>	<ul style="list-style-type: none"> • A child's ability to learn vocabulary is assessed by having the child name a series of pictures, teaching the child the names for pictures they did not know, then retesting using the same pictures to identify response to teaching (test-teach-retest) • A teacher re-words or explains questions to determine if poor performance is influenced by not understanding assessment questions; or the teacher modifies the task (such as providing extra visual supports) to compare performance under different conditions (testing the limits) • The child's performance on a task is assessed using varied levels of prompting to determine the level or degree of prompting required to learn a skill or successfully complete a task (graded levels of prompting)
<p>Decontextualised – Hierarchical: Naturalness of communication:</p> <ul style="list-style-type: none"> - Discrete or 'pure' skills are assessed, which may be used to infer functional performance. - If conducted by a person: Tasks are directed by the assessor, typically in a 'test' format. - If proxy-reported: Skills, usually skills that the child demonstrates without support, are documented without reference to a specific communicative situation or context. <p>Structure of assessment:</p> <ul style="list-style-type: none"> - Assessment is highly structured. Each question or item follows on from previous questions or items in a hierarchical (usually developmental) order. - Presentation of subsequent tasks or questions often depends on success with earlier tasks. <p>(Koole, Nelson, & Curtis, 2015; Mislevy, Steinberg, & Almond, 2002; Schraeder et al., 1999; Skeat & Perry, 2008; Ukrainetz, 2015b; Westby, 2007)</p>	<ul style="list-style-type: none"> • A parent questionnaire asks about the morphological and syntactic abilities that a child demonstrates. Questions are sequenced in order of developmental acquisition, however do not refer to particular communicative situations e.g. Does the child: speak with 3-4 word sentences; use 'ing' verb endings; use 's' regular plural? • An SLP assesses morphological and syntactic skills in a series of cloze questions with picture stimulus: "This girl is running, this boy is _____", with questions presented in order of developmental acquisition. • A teacher completes a checklist profiling a student's pre-linguistic behaviours at school. Questions are sequenced in developmental order, however do not refer to particular communicative situations e.g. Does the student express pleasure and do they do this through facial expression, body language or gesture?; Does the student request desired items and do they do this through facial expression, body language or gesture?
<p>Decontextualised - Non-Hierarchical: Naturalness of communication: Same as for decontextualised – hierarchical (see above) Structure of assessment:</p> <ul style="list-style-type: none"> - Questions or tasks are presented one at a time in a structured manner, but do not follow a set hierarchy or sequence (questions could be administered in a different order without consequence). - Questions or items are different from previous questions or items (tasks are not clearly identifiable as following-on from each other). <p>(Koole et al., 2015; Mislevy et al., 2002; Schraeder et al., 1999; Skeat & Perry, 2008; Ukrainetz, 2015b; Westby, 2007)</p>	<ul style="list-style-type: none"> • A screening checklist asks about behaviours that may indicate language difficulties. Questions are not related to a particular context and are not presented in defined order or sequence e.g. Does the child often: appear to have difficulty thinking of names of objects?; make grammatical errors when speaking?; have difficulty following instructions with 2-3 steps? • Knowledge of social rules is assessed through a series of questions that are not related to specific situations in which the child communicates and are not presented in a developmental sequence or hierarchy of difficulty e.g. "What might it mean if someone says "Pull-up your socks"?; "What might the doctor say when he greets a patient?"

<p>Contextualised: Naturalness of communication</p> <ul style="list-style-type: none"> - Skills are assessed in a meaningful communicative context. Discrete skills may be targeted, but this occurs within the broader context of a naturalistic communicative situation. - If conducted by a person: Tasks are directed by the assessor but occur in a naturalistic context (e.g. book reading) or a contrived scenario representative of a real situation (e.g. role play). Tasks center on a theme (e.g. a story) with topic continuity across tasks. - If proxy-reported: The child's skills are reported in the context of specific communicative situations or contexts i.e. how does the child communicate in a particular situation. <p>Structure of assessment</p> <ul style="list-style-type: none"> - Task presentation is less structured and does not typically follow a hierarchical or developmental sequence (as the focus is on meaningful interaction) <p>(Koole et al., 2015; Mislevy et al., 2002; Schraeder et al., 1999; Skeat & Perry, 2008; Ukrainetz, 2015b; Westby, 2007)</p>	<ul style="list-style-type: none"> • A parent questionnaire assesses communication for different communicative purposes in relation to specific contexts or situations e.g. <i>what does the child do: if they want a toy that is placed out of reach?; when they need to go to the toilet?; if a parent doesn't understand the message they are trying to communicate?</i> • Syntactical skills are examined from a transcription of the child recounting their recent trip to the zoo (i.e. <i>microstructure analysis</i>). • During interactive book reading activities, an SLP assesses the level of support that a child needs to answer questions. The SLP asks questions about the book using different types of questions and observes the child's response to supports such as repetition of questions and visual prompts • A child's ability to respond appropriately to others is observed whilst role-playing real-life scenarios that may occur at school
<p>Decontextualised – Hierarchical: Naturalness of communication:</p> <ul style="list-style-type: none"> - Discrete or 'pure' skills are assessed, which may be used to infer functional performance. - If conducted by a person: Tasks are directed by the assessor, typically in a 'test' format. - If proxy-reported: Skills, usually skills that the child demonstrates without support, are documented without reference to a specific communicative situation or context. <p>Structure of assessment:</p> <ul style="list-style-type: none"> - Assessment is highly structured. Each question or item follows on from previous questions or items in a hierarchical (usually developmental) order. - Presentation of subsequent tasks or questions often depends on success with earlier tasks. <p>(Koole et al., 2015; Mislevy et al., 2002; Schraeder et al., 1999; Skeat & Perry, 2008; Ukrainetz, 2015b; Westby, 2007)</p>	<ul style="list-style-type: none"> • A parent questionnaire asks about the morphological and syntactic abilities that a child demonstrates. Questions are sequenced in order of developmental acquisition, however do not refer to particular communicative situations e.g. <i>Does the child: speak with 3-4 word sentences; use 'ing' verb endings; use 's' regular plural?</i> • An SLP assesses morphological and syntactic skills in a series of cloze questions with picture stimulus: <i>"This girl is running, this boy is _____"</i>, with questions presented in order of developmental acquisition. • A teacher completes a checklist profiling a student's pre-linguistic behaviours at school. Questions are sequenced in developmental order, however do not refer to particular communicative situations e.g. <i>Does the student express pleasure and do they do this through facial expression, body language or gesture?; Does the student request desired items and do they do this through facial expression, body language or gesture?</i>

References

- American Speech-Language-Hearing Association. (2004). Preferred practice patterns for the profession of speech-language pathology [Position Statement]. Retrieved from www.asha.org/policy
- American Speech and Hearing Association. (1993). Language in brief. Retrieved from <http://www.asha.org/Practice-Portal/Clinical-Topics/Spoken-Language-Disorders/Language-In--Brief/>
- Apel, K. (2014). Clinical scientists improving clinical practices: In thoughts and actions. *Language, Speech, and Hearing Services in Schools*, 45, 104-109. doi:10.1044/2014_LSHSS-14-0003

- Betz, S. K., Eickhoff, J. K., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of Specific Language Impairment. *Language, Speech and Hearing Services in Schools, 44*, 133-143. doi:10.1044/0161-1461(2012/12-0093)
- Beukelman, D. R., & Mirenda, P. (2013). Language development and intervention: challenges, supports and instructional approaches. In *Augmentative and alternative communication: Supporting children and adults with complex communication needs* (pp. 255-278). US: Paul H. Brookes Publishing.
- Bishop, D. V. M., & McDonald, D. (2009). Identifying language impairment in children: combining language test scores with parental report. *International Journal of Language & Communication Disorders, 44*(5), 600-615. doi:10.1080/13682820802259662
- Caesar, L. G., & Kohler, P. D. (2009). Tools clinicians use : A survey of language assessment procedures used by school-based speech-pathologists. *Communication Disorders Quarterly, 30*(4), 226-236. doi:10.1177/1525740108326334
- Dockrell, J. E., & Marshall, C. R. (2015). Measurement issues: Assessing language skills in young children. *Child and Adolescent Mental Health, 20*(2). doi:10.1111/camh.12072
- Eadie, P. (2003). Speech pathology assessment practices: One assessment or many? *Advances in Speech Language Pathology, 5*(1), 65-68. doi:10.1080/14417040510001669081
- Edwards, M., Stredler-Brown, A., & Houston, K. T. (2012). Expanding use of Telepractice in speech-language pathology and audiology *Volta Review, 112*(3), 227-242.
- Hegde, M. N., & Pomaville, F. (2017). Assessment of communication disorders in children. In *Assessment of communication disorders in children*. US: Plural Publishing.
- Hyter, Y. D. (2003). Language intervention for children with emotional or behavioral disorders. *Journal of the Council for Children with Behavioral Disorders, 29*(1), 65-76. doi:10.1177/019874290302900104
- Kamhi, A. G., Masterson, J. J., & Apel, K. (2007). Assessment of preschool and early school-age children with developmental language disorders. In *Clinical decision-making in developmental language disorders*. USA: Paul. H. Brooks Publishing Co.
- Kaminski, R. A., Abbott, M., Aguayo, K. B., Latimer, R., & Good, R. H. (2014). The preschool early literacy indicators: Validity and Benchmark Goals. *Topics in Early Childhood Special Education, 34*(2), 71-82. doi:10.1177/0271121414527003
- Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools, 43*(1), 81-96. doi:10.1044/0161-1461(2011/10-0095)
- Koole, H., Nelson, N. W., & Curtis, A. B. (2015). Factors influencing choices of contextualized versus traditional practices with children and adolescents who have traumatic brain injury. *Language, Speech, and Hearing Services in Schools, 46*, 352-361. doi:10.1044/2015_LSHSS-14-0109
- Larson, V. L., & McKinley, N. L. (2007). Procedures for direct assessment. In *Communication Solutions for Older Students* (pp. 209-268). US: Thinking Publications.
- Law, J., Campbell, C., Roulstone, S., Adams, C., & Boyle, J. (2007). Mapping practice onto theory: The speech and language practitioner's construction of receptive language impairment. *International Journal of Language and Communication Disorders, 43*, 245-263.
- Leaders Project. (2013). Understanding Assessment: Assessment Materials- Dynamic vs. Static Assessment. Retrieved from

<http://www.leadersproject.org/2013/03/01/assessment-materials-dynamic-vs-static-assessment/>

- Mashima, P. A., & Doarn, C. R. (2009). Overview of Telehealth Activities in Speech-Language Pathology. *Telemedicine and e-Health*, 14(10), 1101-1117. doi:10.1089/tmj.2008.0080
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19(4), 477-496. doi:10.1191/0265532202lt240oa
- Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working Memory and Specific Language Impairment: An Update on the Relation and Perspectives on Assessment and Treatment. *American Journal of Speech-Language Pathology*, 19, 78-94.
- Nelson, N. W. (1989). Curriculum-Based Language Assessment and Intervention. *Language, Speech, and Hearing Services in Schools*, 20, 170-184. doi:10.1044/0161-1461.2002.170
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149-170. doi:10.1080/09695940701478321
- Ockey, G. J. (2009). Developments and Challenges in the Use of Computer-Based Testing for Assessing Second Language Ability. *The Modern Language Journal*, 93(1), 836-847. doi:10.1111/j.1540-4781.2009.00976.x
- Olswang, L. B., & Bain, B. A. (1996). Assessment information for predicting upcoming change in language production. *Journal of Speech, Language, and Hearing Research*, 39(2), 414-423. doi:10.1044/jshr.3902.414
- Parsons, S., Law, J., & Gascoigne, M. (2005). Teaching receptive vocabulary to children with specific language impairment: a curriculum-based approach. *Child Language Teaching and Therapy*, 21(1), 39-59. doi:10.1191/0265659005ct280oa
- Paul, R., & Norbury, C. F. (2012). Chapter 2: Assessment. In R. Paul & C. F. Norbury (Eds.), *Language Disorders from Infancy through Adolescence: Listening, Speaking, Reading, Writing and Communicating* (4th ed., pp. 22-60). Canada: Mosby Elsevier.
- Peña, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research*, 49(5), 1037-1057. doi:10.1044/1092-4388(2006/074)
- Richards, J. A., Xu, D., Gilkerson, J., Yapanel, U., Gray, S., & Paul, T. (2017). Automated Assessment of Child Vocalization Development Using LENA. *Journal of Speech, Language, and Hearing Research*, 60, 2047-2063. doi:10.1044/2017_JSLHR-L-16-0157
- Schraeder, T. (2008). Assessment, Evaluation and Individualised Education Programs in Schools. In T. Schraeder (Ed.), *A Guide to School Services in Speech-Language Pathology*. San Diego, CA: Plural Publishing.
- Schraeder, T., Quinn, M., Stockman, I. J., & Miller, J. (1999). Authentic Assessment as an Approach to Preschool Speech-Language Screening. *American Journal of Speech-Language Pathology*, 8(3), 195-200. doi:10.1044/1058-0360.0803.195
- Serry, T., Rose, M., & Liamputtong, P. (2008). Oral language predictors for the at-risk reader: A review. *International Journal of Speech-Language Pathology*, 10(6), 392-403. doi:10.1080/17549500802056128
- Singer, B. D., & Bashir, A. S. (1999). What are executive functions and self-regulation and what do they have to do with language-learning disorders? *Language, Speech, and Hearing Services in Schools*, 30(3), 265-273. doi:10.1044/0161-1461.3003.265

- Skeat, J., & Perry, A. (2008). Exploring the implementation and use of outcome measurement in practice: a qualitative study. *International Journal of Language and Communication Disorders*, 43(2), 110-125. doi:10.1080/13682820701449984
- Starling, J., Munro, N., Togher, L., & Arciuli, J. (2012). Training secondary school teachers in instructional language modification techniques to support adolescents with language impairment: A randomized controlled trial. *Language, Speech and Hearing Services in Schools*, 43(4), 475-495.
- Ukrainetz, T. A. (2006). Assisting student's in becoming self-regulated writers. In T. A. Ukrainetz (Ed.), *Contextualised Language Intervention: Scaffolding Pre-K-12 Literacy Achievement* (pp. 570-572). Eau Claire, WI: Thinking Publications.
- Ukrainetz, T. A. (2015a). Contextualised skill intervention framework: The whole and the parts. In T. A. Ukainetz (Ed.), *School-Aged Language Intervention: Evidence-Based Practices*. US, TX: PRO-ED.
- Ukrainetz, T. A. (2015b). The Foundations of language intervention: Theory and research. In T. A. Ukainetz (Ed.), *School-age language intervention: Evidence-based practices*. US, TX: PRO-ED.
- Vaz, S., Cordier, R., Falkmer, M., Ciccarelli, M., Parsons, R., McAuliffe, T., & Falkmer, T. (2015). Should schools expect poor physical and mental health, social adjustment, and participation outcomes in students with disability? *PLoS ONE*, 10(5), e0126630. doi:10.1371/journal.pone.0126630
- Wade, D. T. (2004). Assessment, measurement and data collection tools. *Clinical rehabilitation*, 18, 233-237. doi:10.1191/0269215504cr183ed
- Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010a). Assessment of children's literacy via an internet-based telehealth system. *Telemedicine and e-Health*, 16(5), 564-575. doi:10.1089/tmj.2009.0161
- Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010b). Internet-based telehealth assessment of language using the CELF-4. *Language Speech and Hearing Services in Schools*, 41(445-458). doi:10.1044/0161-1461(2009/08-0131)
- Westby, C. (2007). Application of the ICF in children with language impairments. *Seminars in Speech and Language*, 28(4), 265-272. doi:0.1055/s-2007-986523
- Westerveld, M., & Claessen, M. (2014). Clinician survey of language sampling practices in Australia. *International Journal of Speech-Language Pathology*, 16(3), 242-249. doi:10.3109/17549507.2013.871336
- Williams, C. (2006). Teacher judgements of the language skills of children in the early years of schooling. *Child Language Teaching and Therapy*, 22(2), 135-154. doi:10.1191/0265659006ctQ04oa
- Wixson, K. K., & Valencia, S. W. (2011). Assessment in RTI: What teachers and specialists need to know. *The Reading Teacher*, 64(6), 466-469. doi:10.1598/RT.64.6.13.
- American Speech-Language-Hearing Association. (2004). Preferred practice patterns for the profession of speech-language pathology [Position Statement]. Retrieved from www.asha.org/policy
- American Speech and Hearing Association. (1993). Language in brief. Retrieved from <http://www.asha.org/Practice-Portal/Clinical-Topics/Spoken-Language-Disorders/Language-In--Brief/>
- Apel, K. (2014). Clinical scientists improving clinical practices: In thoughts and actions. *Language, Speech, and Hearing Services in Schools*, 45, 104-109. doi:10.1044/2014_LSHSS-14-0003
- Betz, S. K., Eickhoff, J. K., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of Specific Language Impairment. *Language*,

- Speech and Hearing Services in Schools*, 44, 133-143. doi:10.1044/0161-1461(2012/12-0093)
- Beukelman, D. R., & Mirenda, P. (2013). Language development and intervention: Challenges, supports and instructional approaches. In *Augmentative and alternative communication: Supporting children and adults with complex communication needs* (pp. 255-278). US: Paul H. Brookes Publishing.
- Bishop, D. V. M., & McDonald, D. (2009). Identifying language impairment in children: combining language test scores with parental report. *International Journal of Language & Communication Disorders*, 44(5), 600-615. doi:10.1080/13682820802259662
- Caesar, L. G., & Kohler, P. D. (2009). Tools clinicians use: A survey of language assessment procedures used by school-based speech-pathologists. *Communication Disorders Quarterly*, 30(4), 226-236. doi:10.1177/1525740108326334
- Dockrell, J. E., & Marshall, C. R. (2015). Measurement issues: Assessing language skills in young children. *Child and Adolescent Mental Health*, 20(2). doi:10.1111/camh.12072
- Eadie, P. (2003). Speech pathology assessment practices: One assessment or many? *Advances in Speech Language Pathology*, 5(1), 65-68. doi:10.1080/14417040510001669081
- Edwards, M., Stredler-Brown, A., & Houston, K. T. (2012). Expanding use of telepractice in speech-language pathology and audiology *Volta Review*, 112(3), 227-242.
- Hegde, M. N., & Pomaville, F. (2017). Assessment of communication disorders in children. In *Assessment of communication disorders in children*. US: Plural Publishing.
- Hyter, Y. D. (2003). Language intervention for children with emotional or behavioral disorders. *Journal of the Council for Children with Behavioral Disorders*, 29(1), 65-76. doi:10.1177/019874290302900104
- Kamhi, A. G., Masterson, J. J., & Apel, K. (2007). Assessment of preschool and early school-age children with Developmental Language Disorders. In *Clinical decision-making in Developmental Language Disorders*. USA: Paul. H. Brooks Publishing Co.
- Kaminski, R. A., Abbott, M., Aguayo, K. B., Latimer, R., & Good, R. H. (2014). The preschool early literacy indicators: Validity and benchmark goals. *Topics in Early Childhood Special Education*, 34(2), 71-82. doi:10.1177/0271121414527003
- Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools*, 43(1), 81-96. doi:10.1044/0161-1461(2011/10-0095)
- Koole, H., Nelson, N. W., & Curtis, A. B. (2015). Factors influencing choices of contextualized versus traditional practices with children and adolescents who have traumatic brain injury. *Language, Speech, and Hearing Services in Schools*, 46, 352-361. doi:10.1044/2015_LSHSS-14-0109
- Larson, V. L., & McKinley, N. L. (2007). Procedures for direct assessment. In *Communication solutions for older students* (pp. 209-268). US: Thinking Publications.
- Law, J., Campbell, C., Roulstone, S., Adams, C., & Boyle, J. (2007). Mapping practice onto theory: The speech and language practitioner's construction of receptive language impairment. *International Journal of Language and Communication Disorders*, 43, 245-263.
- Leaders Project. (2013). Understanding assessment: Assessment materials - dynamic vs. static assessment. Retrieved from <http://www.leadersproject.org/2013/03/01/assessment-materials-dynamic-vs-static-assessment/>

- Mashima, P. A., & Doarn, C. R. (2009). Overview of telehealth activities in speech-language pathology. *Telemedicine and e-Health*, *14*(10), 1101-1117. doi:10.1089/tmj.2008.0080
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, *19*(4), 477-496. doi:10.1191/0265532202lt240oa
- Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and Specific Language Impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology*, *19*, 78-94.
- Nelson, N. W. (1989). Curriculum-based language assessment and intervention. *Language, Speech, and Hearing Services in Schools*, *20*, 170-184. doi:10.1044/0161-1461.2002.170
- Newton, P. E. (2007). Clarifying the purposes of educational assessment *Assessment in Education: Principles, Policy & Practice*, *14*(2), 149-170. doi:10.1080/09695940701478321
- Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *The Modern Language Journal*, *93*(1), 836-847. doi:10.1111/j.1540-4781.2009.00976.x
- Olswang, L. B., & Bain, B. A. (1996). Assessment information for predicting upcoming change in language production. *Journal of Speech, Language, and Hearing Research*, *39*(2), 414-423. doi:10.1044/jshr.3902.414
- Parsons, S., Law, J., & Gascoigne, M. (2005). Teaching receptive vocabulary to children with Specific Language Impairment: A curriculum-based approach. *Child Language Teaching and Therapy*, *21*(1), 39-59. doi:10.1191/0265659005ct280oa
- Paul, R., & Norbury, C. F. (2012). Chapter 2: Assessment. In R. Paul & C. F. Norbury (Eds.), *Language disorders from infancy through adolescence: Listening, speaking, reading, writing and communicating* (4th ed., pp. 22-60). Canada: Mosby Elsevier.
- Peña, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research*, *49*(5), 1037-1057. doi:10.1044/1092-4388(2006/074)
- Richards, J. A., Xu, D., Gilkerson, J., Yapanel, U., Gray, S., & Paul, T. (2017). Automated assessment of child vocalization development using LENA. *Journal of Speech, Language, and Hearing Research*, *60*, 2047-2063. doi:10.1044/2017_JSLHR-L-16-0157
- Schraeder, T. (2008). Assessment, evaluation and individualised education programs in schools. In T. Schraeder (Ed.), *A guide to school services in speech-language pathology*. San Diego, CA: Plural Publishing.
- Schraeder, T., Quinn, M., Stockman, I. J., & Miller, J. (1999). Authentic assessment as an approach to preschool speech-language screening. *American Journal of Speech-Language Pathology*, *8*(3), 195-200. doi:10.1044/1058-0360.0803.195
- Serry, T., Rose, M., & Liamputtong, P. (2008). Oral language predictors for the at-risk reader: A review. *International Journal of Speech-Language Pathology*, *10*(6), 392-403. doi:0.1080/17549500802056128
- Singer, B. D., & Bashir, A. S. (1999). What are executive functions and self-regulation and what do they have to do with language-learning disorders? *Language, Speech, and Hearing Services in Schools*, *30*(3), 265-273. doi:10.1044/0161-1461.3003.265

- Skeat, J., & Perry, A. (2008). Exploring the implementation and use of outcome measurement in practice: A qualitative study. *International Journal of Language and Communication Disorders, 43*(2), 110-125. doi:10.1080/13682820701449984
- Starling, J., Munro, N., Togher, L., & Arciuli, J. (2012). Training secondary school teachers in instructional language modification techniques to support adolescents with language impairment: A randomized controlled trial. *Language, Speech and Hearing Services in Schools, 43*(4), 475-495.
- Ukrainetz, T. A. (2006). Assisting student's in becoming self-regulated writers. In T. A. Ukrainetz (Ed.), *Contextualised language intervention: Scaffolding Pre-K-12 literacy achievement* (pp. 570-572). Eau Claire, WI: Thinking Publications.
- Ukrainetz, T. A. (2015a). Contextualised skill intervention framework: The whole and the parts. In T. A. Ukrainetz (Ed.), *School-aged language intervention: Evidence-based practices*. US, TX: PRO-ED.
- Ukrainetz, T. A. (2015b). The foundations of language intervention: Theory and research. In T. A. Ukrainetz (Ed.), *School-age language intervention: Evidence-based practices*. US, TX: PRO-ED.
- Vaz, S., Cordier, R., Falkmer, M., Ciccarelli, M., Parsons, R., McAuliffe, T., & Falkmer, T. (2015). Should schools expect poor physical and mental health, social adjustment, and participation outcomes in students with disability? *PLoS ONE, 10*(5), e0126630. doi:10.1371/journal.pone.0126630
- Wade, D. T. (2004). Assessment, measurement and data collection tools. *Clinical Rehabilitation, 18*, 233-237. doi:10.1191/0269215504cr183ed
- Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010a). Assessment of children's literacy via an internet-based telehealth system. *Telemedicine and e-Health, 16*(5), 564-575. doi:10.1089/tmj.2009.0161
- Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010b). Internet-based telehealth assessment of language using the CELF-4. *Language Speech and Hearing Services in Schools, 41*(445-458). doi:10.1044/0161-1461(2009/08-0131)
- Westby, C. (2007). Application of the ICF in children with language impairments. *Seminars in Speech and Language, 28*(4), 265-272. doi:0.1055/s-2007-986523
- Westerveld, M., & Claessen, M. (2014). Clinician survey of language sampling practices in Australia. *International Journal of Speech-Language Pathology, 16*(3), 242-249. doi:10.3109/17549507.2013.871336
- Williams, C. (2006). Teacher judgements of the language skills of children in the early years of schooling. *Child Language Teaching and Therapy, 22*(2), 135-154. doi:10.1191/0265659006ctQ04oa
- Wixson, K. K., & Valencia, S. W. (2011). Assessment in RTI: What teachers and specialists need to know. *The Reading Teacher, 64*(6), 466-469. doi:10.1598/RT.64.6.13

**Describing Language Assessments and Interventions: A Delphi Study
Supplementary Materials II.**

Table summarising the qualitative and quantitative data collected in each Delphi study round and the associated changes to the taxonomy as a result of data analysis

Qualitative data: Themes from comments	Qualitative data: Examples of participant comments related to the identified themes	Quantitative data: Level of agreement	Changes implemented after each round:
	R1 = Comment from Round one R2 = Comment from Round two R3 = Comment from Round three NA = Not applicable for this round (as no comments were made related to this theme)	R1 :Round one R2: Round two R3: Round three	(Note: no changes after Round three as this was the last round) R1 = Changes after Round one R2 = Changes after Round two NA = Not applicable for this round
Aspect I Suggestion to change sequence in flowchart by placing ‘comprehension’ & ‘production’ after the other domain categories	R1: “Consider if the domains should come before comprehension and production. Much of language requires the integration of comprehension and production so may be better to consider which domain the child is most challenged in before considering receptive versus expressive (if this is even applicable). Not every language domain has a dominant comprehension or production component” R2: NA R3: NA	This suggestion was not linked to lack of consensus but was actioned to improve the taxonomy.	R1: Structural change in aspect I. Components ‘Comprehension’ and ‘Production’ were placed after other domain categories in the taxonomy flowchart. R2: NA
Aspect I Suggestion to add clarification to ensure that categorisation of pre-linguistic communication is clear	R1: “As the taxonomy is valid for school age children regardless of severity etc, potentially an element that incorporates pre-symbolic and pre-intentional ‘spoken language?’ R2: NA R3: NA	This suggestion not linked to lack of consensus but was actioned to improve the taxonomy.	R1: Additional information and examples were added to indicate how assessments targeting pre-linguistic communication may be categorised. R2: NA
Aspect I Identification of overlap between categories of ‘Discourse’ and ‘Social Abilities’	R1: NA R2: “I agree with some definitions for the Domains. I do not agree that Discourse only relates to the types listed, as conversation is a type of discourse, so much of what is classified as ‘social abilities’ is an aspect of ‘Discourse’” R3: NA	R1 and R2: Many participants selected both (or neither) ‘Discourse’ and ‘Social Abilities’ when describing assessments, indicating potential problems with overlap between these categories.	R1: Additional information was added to the definition of ‘Discourse’ and ‘Social Abilities’ to create greater distinction between these two categories. R2: Amalgamation of ‘Discourse’ and ‘Social Abilities’ into one category.
Aspect I Identification that participants may be considering other possible ways an assessment could be conducted, rather than describing assessments as they were used in case studies	R1: NA R2: NA R3: “People may choose semantics as through language sampling you can calculate TTR [<i>type token ratio</i>] and NDW [<i>number of different words</i>], however, your case study did not outline this as an analysis used”	R1-R3: Lack of consensus on application of components ‘semantics’ and ‘executive functioning’.	R1: NA R2: NA
Aspect I Identification that participants may be describing all possible domains, rather than key domains being targeted by the assessment	R1: “The CELF-4 utilises meta-linguistic skills in the items, though it is not explicitly tested. Working memory is also assessed but I wouldn’t classify the CELF4 as assessing broader executive function, and the ability to sustain attention is qualitative data obtained from the assessment process” R2: NA R3: “Possibly clinicians thinking more about the secondary skills involved in the questions in the case study e.g. to initiate a conversation with others you need to use semantic skills, but there is also an element of forward planning. This I would say is a ‘secondary’ skill	R1-R3: Lack of consensus on application of components ‘semantics’, and ‘executive functioning’.	R1: Add additional clarification to highlight that domains are only selected if they are specifically targeted and measured by an assessment. R2: Reduce options for this aspect in the survey (participants may only select one other category in addition to categories that reached consensus in round 2) to determine if consensus is reached on a main domains.

Qualitative data: Themes from comments	Qualitative data: Examples of participant comments related to the identified themes R1 = Comment from Round one R2 = Comment from Round two R3 = Comment from Round three NA = Not applicable for this round (as no comments were made related to this theme)	Quantitative data: Level of agreement R1 :Round one R2: Round two R3: Round three	Changes implemented after each round: (Note: no changes after Round three as this was the last round) R1 = Changes after Round one R2 = Changes after Round two NA = Not applicable for this round
	tapped into indirectly - some clinicians might think that the taxonomy factors in these 'secondary' skills"		
Aspect I Identification of possible overlap between categories 'semantics' and 'executive functioning' with other categories	R1: NA R2: NA R3: "It is hard to separate the categories of 'semantics' and 'executive functioning' out as with a case like this as they would likely influence each other".	R1-R3: Lack of consensus on application of components 'semantics' and 'executive functioning'	R1: NA R2: NA
Aspect I Identification that applying the taxonomy to describing case studies may require a high level of information processing.	R1: NA R2: NA R3: "The amount of information needed to be taken into account in the case studies [<i>may influence application</i>]"	R1-R3: Lack of consensus on application of some components of the taxonomy.	R1: NA R2: NA
Aspect II Lack of clarity with 'prognostic' categories, particularly the 'predict outcome' category.	R1: "I am not sure of any [<i>assessments</i>] in the predict outcome or plan dosage categories" R2: "'Prognostic' tends to lead the reader to the question of whether the young person is likely to improve with or without intervention. 'Predict outcome' then tends to make the reader think about this too rather than about supports the young person would need". R3: "'Predict outcome' is not always 'intuitive' to the definition"	R1-R3: Lack of consensus with selection of 'prognostic'" components to describe assessments	R1: Examples added to show how 'prognostic' categories apply to describing assessments. R2: Examples revised to further highlight application of categories, particularly 'predict outcome' category. A name change for category 'predict outcome' was considered, but not implemented due to inability to identify a more suitable name.
Aspect II Identification that descriptions of purpose of assessment by be influenced by contextual factors related to service policy (e.g. service policy may assign dosage based on diagnosis rather than response to intervention).	R1 "...the concept of 'dosage' is commonly influenced by many other factors (service restraints, funding, availability)..." R2: "I would agree 'specific purpose' section of the assessment purpose, however would rarely separate the prognostic and analytic areas. Assessment usually requires both areas to be covered at the same time in order to meet the reporting and educational requirements on the service" R3: NA	R1-R3: Lack of consensus with application of aspect II categories to describe assessments	R1: The assessments being categorised in the Delphi study were placed in into case studies to provide context R2: Participants were instructed to categorise the assessments in the Delphi study according to the reason for selection in the case study and as though service policy is not an influence
Aspect II Identification that purpose of assessment may be influenced by SLT perspective (e.g. an assessment that is not typically considered diagnostic may be used by SLTs in this way; or if SLT views outcome only as change in diagnostic status, then they may identify 'detect change' as being the same as 'diagnostic').	R1: "Categorising in this area becomes difficult as the waters easily become muddled between the purpose of the tools (intent/design of the tool) and purpose of use (intent of the examiner). Typical purpose may vary according to clinical context and SLT role" R2: [<i>Aspect II</i>] is particularly challenging to categorise, as often this has to do with the nature of the data uncovered and the intent of the clinician in this case. R3: "Perhaps 'diagnostic' because some comparison may be made with peers in the mind of the SLT, though the tool as such doesn't make the comparison"	R1-R3: Lack of consensus with selection of aspect II categorises to describe assessments	R1: The assessments being categorised in the Delphi study were placed in into case studies to provide context. R2: Participants were instructed to categorise the assessments in the Delphi study according to the reason for selection in the case study and as though service policy is not an influence.
Aspect II Identification that lack of consensus may arise if participants are considering all possible ways a tool could be used, rather than categorising based only on how assessment is used in the	R1: NA R2: NA R3: "Conversation & narrative samples are often analysed using [the Systematic Analysis of Language Transcription (SALT) database] which does allow for comparison to peers. Some clinician's may have assumed that [SALT was being used], therefore choosing 'diagnostic'".	R1-R3: Lack of consensus with selection of aspect II categorises to describe assessments	R1: NA R2: NA

case study.			
Aspect III Lack of clarity with term 'Internet'.	R1: "Examples of internet based are not all using the 'internet' so a possibly confusing term to use if covering other than 'internet'. Would technology or {Information and Communication technologies (ICT)] be better?" R2: NA R3: NA	R1: Lack of consensus with identification of assessments as being able to be conducted via telehealth.	R1: Change term 'internet' to 'ICT'. R2: NA
Aspect III Lack of clarity with structure of aspect III.	R1: "...if you have two areas - delivery and setting why you don't have a box with these labelled in between the Aspect III box & the 8 boxes divided into the 2 categories?" R2: NA R3: NA	R1-R3: Lack of consensus across Aspect III	R1: Change the structure of Aspect III to show a component for 'Method' and a component for 'Environmental Context'. R2: NA
Aspect III Identification that lack of consensus may arise from differences between purposes for which assessments are used due to differences in SLT perspective.	R1: "These responses reflect my use of the CELF-4 only and do not necessarily encompass how else the test may be delivered" R2: NA R3: NA	R1-R3: Lack of consensus across Aspect III	R1: The assessments being categorised in the Delphi study were placed in into case studies to provide context. R2: NA
Aspect III Lack of clarity with definition of 'Software'.	R1: "Computer programs and Apps play an important role in language sample analysis, but do not deliver the assessment, as such. Similarly, the CCC-2 can be scored using software, but is not delivered in this way" R2: NA R3: NA	This suggestion not linked to lack of consensus	R1: Clarify that the term 'software' only applies when the assessment is primarily delivered by a software program. R2: NA
Aspect III Lack of clarity with definitions for 'environmental context' with some participants interpreting this as being physical location, rather than 'environmental context'.	R1: "Assessment may be conducted in the clinic or school but draw on child performance in another setting such as home or community. The definitions may then be unclear/confusing" R2: "Difficulty in relation to [case study two] and describing environment. Seen at school but in a withdrawal situation which more closely resembles clinic than classroom environment" R3: "Perhaps its due to an intuitive level of response - as the interview was conducted in the clinic although [it] is a proxy report. Maybe it's just hard to tick home when the interview is in the clinic?"	R1-R3: Lack of consensus with selection of 'environmental context categories to describe assessments.	R1: Change category name from setting' to 'environmental context' to highlight that context is being identified (not physical location). Clarification and examples added to definition to highlight that category identifies 'environment context' and not physical location. R2: Further clarification added to highlight that the category identifies 'environment context' and not physical location.
Aspect III Lack of clarity with definitions for 'school context' with some participants focussing on one element in the assessment, rather than categorising based on the category that best describes the assessment overall.	R1: NA R2: NA R3: "While the assessment is conducted at school it is in a withdrawal/clinical setting. The fact that part of the protocol is that the student brings a piece of school work to share and discuss in the conversational element may lead to confusion."	R1-R3: Lack of consensus with identification of assessments in 'environmental context'	R1: NA R2: NA
Aspect III Lack of clarity with 'environmental context' with some participants confusing the aspect III distinction with 'environmental context' with the Aspect IV distinction 'task type'.	R1: NA R2: "... 'clinical' assessment might be better described as 'de-contextualised' (i.e .focus is on the within-person skills assessed separate from partners and environment where communication occurs) and community might be better described as 'contextualised' (i.e. focus is on the within-person skills assessed within naturalistic interactions with partners in the environment where communication occurs)"	R1-R3: Lack of consensus with identification of assessments in 'environmental context'	R1: NA R2: Further clarification added to definitions to highlight that Aspect III 'environmental context' identifies the environment in which skills are being assessed and Aspect IV 'task type' identifies the communicative tasks used in the assessment.
Aspect III Lack of clarity between 'proxy-reported' vs 'conducted by SLP' with some participants confusing SLT actions (e.g. interviewing a parent) with method by which data is collected (e.g. parent reports information).	R1: "I found the terms 'indirect' and 'reported' were confusing" R2: "Could a 'proxy report' still be recorded in the moment? e.g.: behavioural observation writing down exactly what occurs & this is then reviewed at a later date?" [Note: The behavioural observation described by this participant would be considered assessment conducted by a person and not information obtained through "proxy-report"] R3: NA	R1-R2: Lack of consensus with identification of parent interview/questionnaires as 'proxy-reported' or 'SLT conducted'	R1: Removal of terms 'direct' and 'indirect' from category names. Restructuring of categories within Aspect III to better represent distinctions between categories R2: Clarification added to highlight the difference between 'proxy-reported' and 'conducted by SLP'
Aspect III	R1: NA	R1-R3: Lack of	R1: NA

<p>Identification that applying the taxonomy to describing case studies may require a high level of information processing.</p>	<p>R2: NA R3: "Participants have not read the definitions (and associated examples) properly"</p>	<p>consensus with identification of assessments across Aspect III</p>	<p>R2: NA</p>
---	---	---	---------------

Aspect IV Identification that lack of consensus may arise if participants are considering all possible ways an assessment could be used, rather than categorising based only on how assessment is used in the case study.	R1: "The language sampling protocol can be norm-referenced but only if there is a relevant/appropriate database" R2: "I found 'descriptive' tricky [to identify] with reference to the narrative assessment. They are and can be criterion referenced as well" R3: SLTs might not be familiar enough with the language sampling protocol to know that it is somewhat standardised - often narrative & conversation samples are thought of (& conducted) in a less structured way	R1-R3: Lack of consensus with selection of Aspect IV categories to describe assessments.	R1: The assessments being categorised in the Delphi study were placed in into case studies to provide context. Clarification added to highlight that categories are selected based on how assessments are used in case studies. R2: Further highlight that categories are selected based on how assessments are used in case studies.
Aspect IV Identification that 'task-type' categories 'contextualised' and 'activity-focussed' may be difficult to distinguish between.	R1: "In theory, the definitions were clear, however I found the checklists more challenging to rate based on the definitions between 'contextualised' and 'activity focussed'" R2: "Decision making regarding 'contextualised' and 'activity-focussed' [is] not always clear." R3: "Contextualised and 'activity based' categories overlap to an extent"	R1-R3: Lack of consensus with selection of Aspect IV 'Task Type' categories to describe assessments.	R1: Definitions revised and examples added to assist with distinctions between 'task type' categories. R2: Information on the taxonomy was formatted under headings to assist with application of terms.
Aspect IV Identification that lack of consensus may arise if SLTs apply definitions that are different to definitions in the taxonomy.	R1: NA R2: NA R3: "Possibly [confusion] in regards to my understanding of dynamic assessment? It seems clear in your definition however"	R1-R3: Lack of consensus with identification assessments in case study two as 'standardised' and 'dynamic'	R1: NA R2: NA
Aspect IV Identification that applying the taxonomy to describing case studies may require a high level of information processing.	R1: NA R2: NA R3: "The definitions contain a lot of detail which is hard to hold on to when flipping back [through the reference document] to think about what was done in the assessment"	R1-R3: Lack of consensus with selection of Aspect IV 'Task Type' categories to describe assessments.	R1: NA R2: NA
Overall Taxonomy Participants identified as finding the taxonomy useful for conceptualising clinical work	R1: "I really like this classification. I use most if not all types of assessment but had never considered the different types so explicitly. I think it will add hugely to professional education at [universities] and work places to help build a more conscious and explicit awareness of what we do." R3: "I think it's a great classification and useful"	NA	NA
Overall Taxonomy Participants identified that understanding and applying the taxonomy accurately takes time and consideration	R2: "Challenging to keep all parameters in mind. I hope I have not been too hasty in my responses." R3: "I think the assessment type classification is complex and a new way of thinking. [It] takes real consideration to use."	NA	NA
Overall taxonomy Participants commented that the taxonomy and their understanding of the taxonomy improved over rounds and that examples assisted in improving the taxonomy.	R2: "The definitions were helpful in considering the options." R3: "The new additions to definitions and examples have helped clarify the taxonomy."	NA	NA